

Dolly's legacy

Ten years on, mammalian cloning is moving forward with central societal issues remaining unresolved. Yet human reproductive cloning seems inevitable.

Ten years ago, given a week's notice by *Nature's* usual embargoed press release, journalists were gearing up for what many of them recognized as the hottest science story for years. The press release announced the cloning of sheep from adult cells. A journalist on a British Sunday newspaper, *The Observer*, picked up the same story from a film production company and ran with it ahead of the embargo date. Furious editors and writers on other newspapers scrambled to catch up. "It is the prospect of cloning people, creating armies of dictators, that will attract most attention," said *The Observer*. And so it proved. Within days the president of the United States, the head of the European Commission, the Vatican and many others were calling for a review of the regulations on cloning research, if not an outright ban.

The world was simply not prepared for the debate. The cloning a year previously of sheep from embryonic cells had led Davor Solter, in an accompanying News & Views article in *Nature*, to warn that "it might be a good idea to start thinking about how we might use" the ability to clone from adult cells. But others were dismissive of the prospect, or predicted it to be many years away.

The researchers who cloned Dolly had kept their research under wraps. But learning from their media experiences with embryonic cloning, they had hired a public-relations company and collaborated on a television documentary to be broadcast after publication. Their approach misfired, but their intentions still serve as a model for others: researchers and their institutions have a responsibility to provide their perspective of the context and implications for a broad audience when announcing startling results, both to the media and through their own websites. Journals, including this one, can no doubt do more to help them convey that context.

What of the subsequent research? As we describe on page 800, much of it has focused on the transfer of nuclei of somatic cells in the context

of stem cells. The agricultural implications of Dolly received relatively little public attention at the time, but it is here that reproductive cloning has proceeded apace, with a dozen mammalian species cloned. One measure of that progress is the risk assessment, currently open for public consultation, by the US Food and Drug Administration (FDA; see www.fda.gov/cvm/CloneRiskAssessment.htm).

Although generally upbeat about safety, the FDA report highlights one critical lack of progress: the efficiency of mammalian cloning is still low. The agenda for fundamental biology is clearly highlighted. One area of study, for example, is epigenetics: which chemical markers attached to the DNA are altered? And what exactly happens in chromatin structure, compared with conventional reproduction? We can look forward to progress in these areas. Meanwhile, although US consumers are uneasy about animal cloning, it seems unlikely that they will oppose its application, let alone its products.

In contrast, what has been universally deemed as unacceptable is the pursuit of human reproductive cloning — or the production of what some have called a delayed identical twin. Here, the two issues that have dominated the discussion have been dignity and safety. There is a consensus that dignity is not undermined if a human offspring is valued in its own right and not merely as a means to an end. But there is no consensus that we will eventually know enough about cloning for the risks of creating human clones to be so small as to be ethically acceptable.

The debate may seem to have been pre-empted by prompt prohibition. But as the science of epigenetics and of development inevitably progresses, those for whom cloning is the only means to bypass sterility or genetic disease, say, will increasingly demand its use. Unless there is some unknown fundamental biological obstacle, and given wholly positive ethical motivations, human reproductive cloning is an eventual certainty. ■

Rise to the challenge

The European Research Council, launched next week, is a stimulus for weak universities.

Outside Europe it may be hard to imagine the scale of the triumph involved in the creation of a Europe-wide agency for competitive basic research, along the lines of the US National Science Foundation. After what seems like another hundred-years' war, a solution to member countries' concerns — that they must pay into a pot that then funds researchers elsewhere — has finally been found in the form of the European Research Council (ERC). It launches to great fanfare in Berlin next week. Importantly, it is committed to funding the best science, free of regional and political agendas.

The first call for grant applications will be restricted to young investigators, with a second call, for advanced investigators, to be announced later in the year. But too few European universities are ready to host the recipients.

The grants, which can run for up to five years, will be big: between €100,000 (US\$130,000) and €400,000 per year, and deliberately designed to be prestigious. With a budget of only €300 million this year, competition will be particularly stiff (although annual funds will rise to €1.5 billion by 2013). The two-tier application procedure will be handled by twenty panels of experts, five in social sciences, eight in physical sciences and seven in life sciences. With the advice of specialist referees that they themselves select, the panels will judge the applications on the basis of merit, without reference to the nation involved.

But they will also assess the ability of the host institute to offer an

appropriately supportive environment. This means providing genuine access to good infrastructure and a vigorous intellectual environment — not least to encourage applications for the grant's attendant posts from the best graduates and postdocs.

Researchers applying to the ERC must choose their host institute, and if their home base doesn't offer them much of a package, they can approach any other university or research centre. The phrase "without reference to nation" may at this point begin to seem disingenuous. Some countries are relatively inflexible in the conditions that their universities can offer individual scientists. Their universities may not, for example, be able to offer a salary attractive by local standards if they are hampered by fixed salary scales.

In short, the most flexible universities will be best placed to attract ERC grant holders. This is as it should be. The ERC is in effect a wake-up call for universities to free themselves of their chains and become internationally competitive.

It is fortuitous for Germany that it currently holds the rotating European Union (EU) presidency, and therefore hosts the launch of the ERC — part of the EU's Seventh Framework Research Programme,

which runs until 2013. The German government is currently trying hard to loosen the chains around the country's universities, forged during the 1970s' anti-elitist movement that rigidly imposed equal status on them. Similar events squeezed competition between universities out of other European countries such as France and Italy, which are now also trying to recover. The former communist Central European countries, now members of the EU, have an even longer history of institutionalized academic paralysis.

One of the most effective instruments that Germany has created to re-inject the competitive spirit is the Excellence Initiative, which throws a few million euros and considerable prestige at a handful of universities judged in a high-profile competition to be strongest in research. All universities have been energized in the process. The ERC, if it works as planned, should provide such a stimulus across Europe, and ever more so as its experience and budget grow.

When the German chancellor Angela Merkel opens the ERC launch next Tuesday, she will at the same time be launching a new phase in European research — but only for those universities that are up to the task. ■

Regulatory fist-fight

A move to wrest control of US federal regulations from government agencies should be opposed.

In an executive order passed last month, the administration of President George W. Bush tweaked the terms of the relationship between government agencies and its own Office of Management and Budget. The changes are subtle and arcane, but significant nevertheless. The administration will now review supporting documents as well as the regulations themselves. Agencies will have to present some additional cost-benefit figures. And the official in charge of coordinating all this from the agency end must now be a presidential appointee. This person will initiate rule-making and be "involved at each stage of the regulatory process".

Because deliberations on regulation are open to public scrutiny only after an agency submits its plans to the president's budget office (the Freedom of Information Act does not apply to deliberative processes within agencies), they can be smothered at birth inside the agency by the presidential appointee, away from public scrutiny.

Administration officials have downplayed the significance of these changes and, according to the Congressional Research Service, most of these officials are already presidential appointees. But the move represents yet another incremental power shift. The Bush administration's approach has been to make small bureaucratic changes or insertions here and there that make it more laborious to pass regulations, and easier for industry and the president to have regulations shift in their preferred directions. The influence of well-considered scientific advice has been progressively weakened.

Consider, for example, the Data Quality Act of 2001, which opened the door for industry to take issue with the data used to make regulatory decisions. In 2005, in a move opposed by scientists, the salt industry used it to challenge the findings of a federally funded study

of sodium and blood pressure (see *Nature* **433**, 671; 2005). Consider also two failed attempts, one in 2003 to control the peer review of science informing regulation, and one last year to bundle all regulations into a centralized risk-assessment process run by the budget office. US scientists have the National Academies to thank for fending these off (see *Nature* **442**, 223–224; 2006).

At a hearing last week of the House science committee, Sally Katzen, who ran the department dealing with regulation at President Bill Clinton's budget office, described the effect this way: "Each step has placed a thumb on the scales, and now we have a whole fist."

The fact that this hearing and another in the judiciary committee were held at all is good news. Democrats and Republicans alike should see these moves for what they are: attempts to influence regulations at agencies that have been given their missions by Congress. It is all of a piece with Bush's habit of signing laws with attached statements indicating which bits of the law he doesn't intend to follow.

Congresswoman Linda Sánchez (Democrat, California), chair of the Subcommittee on Commercial and Administrative Law in the judiciary committee, intends to ask the Office of Management and Budget for more information on how the new executive order is to be implemented in practice. But only time will tell whether its provisions have a large or small effect. It is difficult for Congress to overturn an executive order. They do so by passing a law that contradicts it, but this law could be vetoed by the president. It would be better if Congress, encouraged by scientists, were to make such a fuss that the administration backs off.

If no one protests, this order may well be followed by other such manoeuvres, each designed to make science a mere vestigial irritant to the otherwise smooth implementation of Bush's personal will. This would be a bad idea even if the president were a fan of precautionary regulation based on empirical science. But he isn't. ■

"The influence of well-considered scientific advice has been progressively weakened."

RESEARCH HIGHLIGHTS

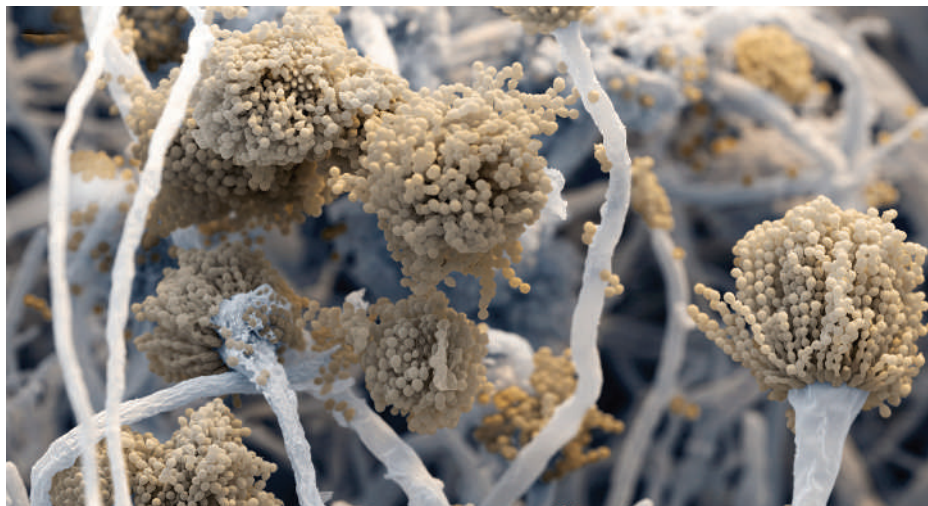
All wrapped up

Cell. Microbiol. doi:10.1111/j.1462-5822.2007.00895.x (2007)

A fungus that causes life-threatening lung infections in people protects itself within a biofilm, researchers at the Pasteur Institute in Paris, France, have found.

Aspergillus fumigatus (pictured) has previously been grown in liquid for study, but Anne Beauvais and her colleagues decided to grow it on a gel exposed to air, to mimic conditions in the lung more closely. They observed that a hydrophobic sheath glued its filaments together.

The existence of this protective film may help to explain why some drugs that have killed the fungus in liquid-based *in vitro* experiments do not always work very well when given to patients.



EYE OF SCIENCE/SPL

CELL BIOLOGY

Torn asunder

J. Cell Biol. doi:10.1083/jcb.200609014 (2007)

A key regulatory protein that sits in animal cell membranes is activated when a neighbouring cell tears a part of it away, new research shows.

The protein, called Notch, is a two-protein complex, part of which is exposed on the cell surface. Removal of the entire extracellular portion frees the intracellular portion of the protein to travel to the nucleus and regulate gene expression.

Gerry Weinmaster and her colleagues at the University of California, Los Angeles, suspect that when a neighbouring cell rips away the exposed part of Notch, it leaves the remaining portion vulnerable to degradation by proteases. Their action could then liberate the inner Notch tail.

STEM CELLS

Personal sperm bank

Dev. Cell **12**, 195–206 (2007)

Mice may have a backup system for producing sperm, report Shosei Yoshida and his co-workers at Kyoto University, Japan. The mechanism may help to ensure that males can make sperm for their entire lives.

The group engineered mice so that undifferentiated sperm cells expressed a fluorescent tag. By observing how the number and location of fluorescent cells changed over time, the group found that mice seem to make sperm from two kinds of cell: 'actual' stem cells, which give rise to sperm-generating cells and renew themselves; and 'potential' stem cells. Potential stem cells normally become sperm without renewing themselves,

meaning their fluorescence switches off. When the actual stem cells fail, however, the fluorescence of potential stem cells persists, suggesting that they have become stem cells themselves.

ARCHAEOLOGY

Spicing it up

Science **315**, 986–988 (2007)

Ancient humans domesticated chilli peppers more than 6,100 years ago, according to newly gathered archaeological evidence. In some regions, this means that chilli cultivation even pre-dates pottery.

The chilli is a relatively recent addition to cuisine in most of the world, however. Before Columbus's travels in the fifteenth century, the spicy peppers of the genus *Capsicum* were limited to the Americas.

To trace the chilli's history in the Americas, anthropologist Linda Perry of the Smithsonian National Museum of Natural History in Washington DC and her colleagues analysed starch grains preserved on cookware and stone tools from several archaeological sites. They found some grains identical to those produced by modern domesticated peppers — showing that ancient humans not only cultivated staple crops such as maize and yam, but also farmed for flavour.



QUANTUM PHYSICS

Not so firm

Nature Physics doi:10.1038/nphys539 (2007)

New doubts about whether physicists have seen evidence for 'supersolid' flow are raised by Philip Anderson of Princeton University, New Jersey.

In a theoretical paper, he suggests that experiments involving a rotating block of helium-4 did not detect supersolidity — a quantum effect whereby atoms in a solid move en masse without viscosity — as claimed, but rather measured the influence of a novel form of vortex liquid.

Vortex liquids occur when whirlpools pepper the quantum wavefunction of a material and are quite well studied in some systems. Anderson suggests that the type of vortex liquid that he predicts might also explain certain puzzling features of high-temperature superconductors.

MATERIALS SCIENCE

Heavyweight lightweight

Adv. Mater. doi:10.1002/adma.200601748 (2007)

Carbon nanotubes can be made into a low-density material known as an 'aerogel', report Arjun Yodh and his colleagues at the University of Pennsylvania in Philadelphia.

Such materials are extremely porous, ultra-light and extra-strong — and a carbon nanotube aerogel is good at conducting electricity to boot.

To make the aerogel, the researchers freeze-dried carbon nanotubes out of a liquid suspension, leaving a network of filaments with a high ratio of surface area to volume. The

FOODPIX/PHOTOLIBRARY.COM

properties of this network could be readily manipulated by tweaking the freeze-drying process, or by adding a polymer to reinforce the structure. The resulting material could be useful for making electrodes or in chemical and thermoelectric sensors.

INFECTIOUS DISEASES

HIV's route of entry

Immunity doi:10.1016/j.immuni.2007.01.007 (2007)
Most new HIV infections in women occur as a result of sexual intercourse, so researchers are working to find agents, such as microbicides, that prevent HIV from infecting the body through the vagina.

To pinpoint the virus's targets, Julie McElrath and her colleagues at the Fred Hutchinson Cancer Research Center in Seattle, Washington, performed experiments on live vaginal tissue removed from women during surgical repair procedures and hysterectomies. They show that two types of immune cell in the vagina — CD4⁺ T cells and Langerhans cells — are vulnerable to attack by HIV, indicating that microbicides and vaccines will have to block the virus from entering both these cell types to be effective.

ORGANIC CHEMISTRY

Better by design

J. Am. Chem. Soc. doi:10.1021/ja067870m (2007)
Most chemists would be happy just to show that they can synthesize some naturally occurring compound, but Kazunori Koide and his colleagues at the University of Pittsburgh, Pennsylvania, have gone one better. Having completed a concise synthesis of FR901464, a compound that shows anticancer activity *in vitro* and in mice, they then prepared a more potent analogue.

Their approach used a reaction known

as cross metathesis, which had rarely been exploited near the end of the synthesis of complicated molecules. The gamble worked, but the researchers then found that FR901464 degrades quite quickly. To address this issue, they prepared a version of the compound in which a hydroxy group (OH) was replaced with a methyl group (CH₃).

This analogue was more stable than the original and was 100-fold more active in cancer cell antiproliferation assays.

ECOLOGY

Pest control

Proc. R. Soc. Lond. B doi:10.1098/rspb.2006.0415 (2007)
Devil's gardens in the Amazon are tracts of forest dominated by just one species of tree, *Duroia hirsuta*. Local legend holds that the stands are tended by evil spirits, but Megan Frederickson and her colleagues have previously shown the gardeners to be ants (*Nature* **437**, 495–496; 2005), which dwell in the tree stems and kill other plants.

Now Frederickson and Deborah Gordon, both of Stanford University in California, ask what stops the gardens, which can include as many as 600 trees spread across 1,000 square metres, from growing even bigger. They suggest that the advantage conferred by the ants is offset by the concentration of leaf-eating creatures in the gardens.

By painstakingly photographing the leaves of *D. hirsuta* trees inside and outside the gardens, they estimate that trees inside gardens lose around three times as much leaf area per year as those outside.

INSECT BEHAVIOUR

Fly by eye

Curr. Biol. doi:10.1016/j.cub.2006.12.032 (2007)
A miniature helicopter that flies using data from just two camera pixels can mimic the way insects pilot themselves in flight, say researchers at the Biorobotics Laboratory of France's Centre National de la Recherche Scientifique in Marseille.

Nicolas Franceschini and his colleagues have designed their helicopter (pictured) to navigate by monitoring the apparent motion



of objects on the ground as it flies over them. The idea that insects might rely on such 'optic flow' to control their flight is fairly well accepted, but Franceschini's team has put the concept on a firmer footing. The researchers present a control scheme that can reproduce a range of insect behaviours.

In this scheme, when the helicopter changes its ground speed, it keeps the optic flow constant by adjusting its altitude. The simple model can explain why insects descend in a headwind and take a straight path towards the ground to land.

H. RAGUET/ CNRS PHOTO THÉQUE

JOURNAL CLUB

Harold Tobin
University of Wisconsin-Madison, USA

A geophysicist wonders how and why faults behave in so many different ways.

I'm involved, with colleagues, in a project of the Integrated Ocean Drilling Program (IODP) to drill deep into the Nankai Trough subduction zone off southwestern Japan — a site of numerous great earthquakes and tsunamis.

Major unknowns in the generation of tsunamis include how far earthquake fault slip can propagate up towards the sea bed and what factors control how that slip stops in accretionary wedges — the submarine mountain ranges created as sediment and rock are scraped off the sinking plate.

My research focus is on faults in such wedges, which are generally thought of as aseismic, or incapable of earthquakes. By drilling into the wedge faults at Nankai Trough, we hope to learn how aseismic faults give way

with depth to the seismic faulting associated with tsunamis.

Recently, a new kind of slow-motion earthquake was observed in this wedge (Y. Ito & K. Obara, *Geophys. Res. Lett.* **33**, L02311; 2006). Suddenly, wedges don't seem so aseismic after all.

These 'very-low-frequency' earthquakes, some as large as magnitude 4.4, have previously gone unrecognized because their seismic waves don't show up in the frequency range in which earthquakes are normally detected.

By chance, the quakes were

detected exactly where my IODP team plans to start drilling later this year. We hope to install sensors deep in the subsurface to record the earthquakes up close and to measure pore fluid pressure and strain in the rock. We'll also collect samples for laboratory studies of the frictional properties of the rock.

Taken together, the *in situ* and sample data should yield insight into the processes responsible for these slow-motion quakes. This might help us to understand the aseismic-seismic transition.

SPECIAL REPORT

Dolly: a decade on

Ten years ago, the birth of Dolly the sheep sparked a media frenzy and a prolonged ethical debate. Today, the arguments have switched focus to stem cells, and the research itself is beginning to change tack.

"Scientists clone adult sheep — triumph of UK raises alarm over human use," ran the first headline announcing the cloning of an adult mammal ten years ago this week. Ian Wilmut at the Roslin Institute near Edinburgh and his colleagues at PPL Therapeutics in East Lothian reported on 27 February 1997 that they had produced a lamb named Dolly, born the previous July, that was the first mammalian clone created using the genetic material from an adult cell¹.

As soon as the story hit the front page (the news was broken by a British Sunday newspaper four days ahead of *Nature's* publication), a public and media maelstrom ensued.

"The first press calls came from New Zealand," recalls Sue Charles of Northbank Communications in London, who was handling publicity for Roslin and PPL at the time. "They worked their way through Australia, Asia and Europe." Later in the day, calls began coming in from the United States, with interest ranging from the science involved to politics and religion. "We even had a US chat show that wanted Dolly on," says Charles. "They offered to fly her over." She remembers that her team, together with Wilmut and his colleagues, took around 2,000 calls from journalists in two weeks.

But Dolly wasn't popular with everyone. Pundits warned of a future in which armies of human clones would be created by the evil and egotistical. Conservatives predicted the demise

of the nuclear family. Activists cautioned that fertility doctors would perfect the technique and get rich making clones for the infertile, the narcissistic and the eccentric. US President Bill Clinton announced that the feat "raises serious ethical questions" and commanded his bioethics advisers to report to him in 90 days on measures he should take to prevent its abuse.

Ten years later, the ethical debate launched by Dolly, and encouraged by science-fiction stories, has changed. It has been supplanted by one that is more complex, more rooted in reality and far more relevant to the research that scientists want to do.

"What didn't happen was the birth of a [cloned] child or a widespread public demand for the use of cloning for reproduction," says Alta Charo, a professor of law and bioethics at the University of Wisconsin, Madison, who served as a bioethics adviser to Clinton. "What did happen was a complete shift in the ethical discussion from reproductive uses of cloning to research uses. And a merging of the cloning debate into the debate around embryonic stem-cell research — to the disadvantage of both fields because of the attendant confusion."

Back in early 1997, none of Wilmut and his colleagues, the referees who reviewed their paper, or the *Nature* editors who oversaw it, anticipated the huge public reaction to the cloning of Dolly. Scientists in the field saw her birth as an incremental advance — in large

"This is one of those areas where just trying to rein in nutty behaviour became a full-time job."



Celebrity sheep: Dolly, the first clone of an adult mammal, became an unlikely media star.

part because one year earlier, *Nature* had published a paper from Wilmut's group reporting the cloning of two lambs, Morag and Megan, using nuclei from embryonic cells².

"I always maintained that Dolly was expected and Morag and Megan were truly surprising," says Davor Solter, director of the Max Planck Institute for Immunobiology in Freiburg, Germany. Solter wrote a News & Views article in *Nature* about the paper on Morag and Megan, suggesting that it was time to start thinking about the implications and uses of cloning mammals from adult cells³.

Philip Campbell, *Nature's* editor-in-chief, also recalls that the media storm over Dolly took him by surprise. "Staff and referees were aware that this was the paper that in principle demonstrated how to clone mammals, including humans," he says. "But neither they nor I

D. M. DENNIS/PHOTOLIBRARY; FLPA; ROSLIN INST.

CLONING
TIMELINE

1952

Robert Briggs and Thomas King in Philadelphia, Pennsylvania, describe how they cloned frogs (*Rana pipiens*) by replacing the nuclei of eggs with cells from tadpoles and adult intestinal epithelium. A similar experiment was first proposed by Hans Spemann at the University of Freiburg, Germany, in 1938.



1984

Chinese researchers clone a fish — the crucian carp (*Carassius carassius*) — from cultured kidney cells.



1996

Researchers at the Roslin Institute in Scotland clone two lambs — Megan and Morag — from embryonic cells. This was a crucial step towards cloning an animal from an adult cell, and is seen by some scientists as a bigger breakthrough than Dolly herself.

1997

Roslin researchers announce the birth of Dolly the sheep, the first mammal to be cloned from an adult cell, igniting public debate about the prospects for cloning humans.



**BLOGS TO THE RESCUE!**

Proposal calls for web community to help professionals in disaster relief.

www.nature.com/news

M. POLAK/CORBIS SYGMA

Philadelphia. "This is one of those areas where just trying to rein in nutty behaviour became a full-time job."

But as the decade passed and a menagerie of other mammals was cloned (see 'Cloning timeline'), no cloned human babies appeared. What did occur, and what moved the ethical debate in an unforeseen direction, was the isolation of human embryonic stem-cell lines by James Thomson and his colleagues at the University of Wisconsin, Madison⁴.

With that achievement, it became clear that broad research avenues could be opened up by cloning human embryos to extract stem cells from them. These could then be used as disease models and drug targets, or to develop therapies involving tissue transplantation.

But as quickly as scientists recognized the potential of such opportunities, political and ethical opponents seized on the notion that allowing cloning in research would only ensure that it would one day be used for reproductive purposes. What's more, they argued, research cloning was a fundamental assault on human dignity, because it creates, manipulates and destroys human embryos for scientific ends.

"What really took place is that the stem-cell debate replaced the cloning debate," says Caplan. "Because there was — and is — a tremendous interest in trying to clone embryos, not people."

The South Korean scandal of Seoul National University's Woo Suk Hwang, whose claims to have derived stem-cell lines from cloned human embryos were proved to be fraudulent in 2006, generated plenty of bad press for the field. But there are signs that the new debate is taking a different course from that on reproductive cloning because of the potential of stem-cell research to improve human health.

In the United States, for example, opposition to cloning babies has remained firm through a decade of polling, at about 90%, but polls in recent years have shown that 60–70% of the public supports research using stem cells obtained from discarded embryos in fertility clinics.

"As people learn about the possibilities for new approaches to disease, they see the embryonic stem-cell issue in a different framework," says Jonathan Moreno, a bioethicist at the University of Pennsylvania, who co-chaired a committee that crafted 2005 research guidelines for the US National Academies. "They see it as medical research that could help them or their families."

The notion of cloning embryos to be a source of stem cells — as opposed to using embryos left over from fertility treatment that are slated for destruction anyway — is much more controversial. It remains a touchy political issue in many countries, including the United States, and is approached gingerly by public and private funders alike.

In the near term, "I rather doubt that we will see very much [cloning] in the context of embryonic stem-cell research in the United States," says Moreno. But he thinks advances are likely to come in countries where the work is seen as more acceptable, such as Britain, where groups led by Wilmut at the University of Edinburgh and Alison Murdoch at the International Centre for Life in Newcastle upon Tyne have been given permission to pursue it. (Murdoch's group has already cloned at least one human embryo, but has had no luck extracting stem-cell lines.)

As embryonic stem-cell research allows the genetic contributions to complex disease to be teased out, Charo believes that the ethical debate will shift again. Rather than focusing on the moral implications of cloning itself, she says, questions of an individual's responsibility for their known predisposition to disease will come to the fore: "If we move towards genetics research, we will have a new set of issues that *Time* and *Newsweek* can put on their covers." ■

Meredith Wadman

1. Wilmut, I. *et al. Nature* **385**, 810–813 (1997).
2. Campbell, K. H. S. *et al. Nature* **380**, 64–66 (1996).
3. Solter, D. *Nature* **380**, 24–25 (1996).
4. Thomson, J. A. *et al. Science* **282**, 1145–1147 (1998).

See page 802 and Editorial, page 795.

anticipated the furore. We were very focused on just how much of an advance it represented on Morag and Megan a year earlier."

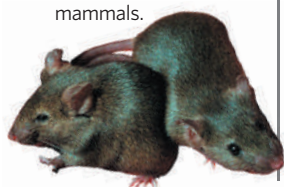
But what to scientists was one small step, built on work dating back decades, was to the public the herald of a brave — and unwelcome — new world in which human clones would become as common as test-tube babies.

The claims that soon followed (of intentions to clone babies, or of having accomplished the deed) from sources including Chicago physicist Richard Seed, Italian embryologist Severino Antinori and the Raelian cult, which believes that the human race was cloned from aliens, only fanned the flames of the public imagination. At the outset, "kooks, cultists and commoners made all the news about cloning," recalls Arthur Caplan, director of the Center for Bioethics at the University of Pennsylvania in

S. MOITESSIER/AP; Y. TSUNODA/AGRESEARCH; REUTERS/CORBIS

1998

Scientists at the University of Hawaii reveal the cloning of three generations of mice from the nuclei of adult cells, suggesting the technique could work on other mammals.



1998

Japanese researchers report cloning eight calves using adult cells from slaughterhouse entrails, raising the possibility that animals could be cloned for the quality of their meat.



1998

Scientists in New Zealand announce Elsie, a clone created from an adult cell from the last surviving Enderby Island cow (*Bos gaurus*). Attempts to clone endangered species have met with criticism that the technique will do little good without concurrent habitat preservation.



2000

PPL Therapeutics in Scotland unveils a litter of five cloned piglets. The firm says that genetically engineered cloned pigs could one day provide a source of organ transplants for humans.



2002

The first cloned cat (*Felis domesticus*), named cc for 'copycat', is announced by Texas A&M researchers. Cc's coat pattern is not the same as her genetic donor's, showing the impact on development of non-genetic effects.



Dolly: a hard act to follow

Four years after Dolly's death, scientists are still amazed she was ever born. That's because the decade since the announcement of Dolly's birth has brought cloning researchers few triumphs and many frustrations. They may have learned a little more about the events that allow an egg to reprogramme the nucleus of an adult cell to produce a new organism. But these advances haven't led to big improvements in the cloning process, or yielded huge commercial payoffs.

In many ways, scientists are still fumbling in the dark, says Robert Lanza, vice-president for research and development at Advanced Cell Technology in Worcester, Massachusetts. "When Dolly was born, we thought that in a few years we would understand the magic in the egg that allows it to reprogramme a cell's DNA," Lanza says. "But cloning is still essentially a black box."

Frogs were cloned as early as 1952, but mammals proved too problematic until scientists at the Roslin Institute in Edinburgh cloned two sheep — Megan and Morag — from embryonic cells in 1995. But it was the group's creation of Dolly, cloned from an adult udder cell, that finally overturned the idea that in mammals, developed cells could not reverse their fate.

After Dolly, scientists thought they would soon be able to clone many other mammals. Some envisaged an industry of cloning applications, from the production of medicines in live bioreactors — cloned, genetically modified livestock — to the creation of herds of cloned food animals.

Scientists have made some progress towards potential applications. For instance, they have cloned transgenic pigs that might one day be used as organ donors, transgenic cows and goats that produce proteins that some humans lack — they have even cloned pets. But the low

efficiency of the cloning process has stymied industrial development, and most companies set up to commercialize cloning have shut down.

Only 2–5% of cloned animal embryos grow into healthy offspring. This is slightly better than a decade ago — Dolly was the only lamb born from 277 cloned embryos — but it is still far below the efficiency demanded by industry.

There are several reasons why it has been difficult to figure out how cloning works. Early development is slightly different in every species, so cloning a new species is an arduous, trial-and-error process that requires scientists to use thousands of eggs. But eggs aren't always easy to obtain, especially for monkeys and humans, and no clones of either have been born. And even if eggs are available, some of scientists' traditional model species, such as mice, have proved extremely problematic.

Yet some see a way ahead. Many of the abnormalities that doom cloned embryos are due to faulty epigenetics — a set of controls that silences or activates genes by chemically modifying the DNA or by binding it to certain proteins. Epigenetic processes guide young cells with limitless potential into restricted adult fates, by shutting down certain genes. During cloning, an adult nucleus is transplanted into an egg, which must then erase the adult genome's epigenetic marks, so it can re-express every gene necessary to build a new animal.

But requiring an egg to do this reprogramming on its own bypasses other natural mechanisms that normally help reset a cell's epigenetic programme. Perhaps it's not surprising that the egg doesn't always do a perfect job, says Ian Wilmut, Dolly's creator and director of the Scottish Center for Regenerative Medicine at the University of Edinburgh. "When you think about what we're asking the egg to do for us, in a way, I think we should still be surprised

that cloning works at all," Wilmut says.

Rudolf Jaenisch, a biologist at the Massachusetts Institute of Technology's Whitehead Institute for Biomedical Research, is among those trying to decipher these epigenetic glitches. If they can understand what goes wrong with the reprogramming, they may be able to fix it — or, better yet, initiate the process themselves by adding chemicals or proteins to adult cell nuclei. This would allow scientists to bypass the need for eggs altogether.

The field is moving in this direction, trying to find ways to reprogramme cells without using eggs. Last summer, for instance, Shinya Yamanaka and Kazutoshi Takahashi from Kyoto University in Japan used a cocktail of proteins to turn adult mouse cells into more flexible cells that behave a lot like stem cells (K. Takahashi and S. Yamanaka *Cell* 126, 663–676; 2006).

This molecular approach solves the efficiency problem and avoids society's ethical qualms about cloning (see page 800). "For many of us, that is the holy grail, and nuclear transplantation is an intermediate stage," says Alan Colman, chief executive of Singapore-based company ES Cell International and former research director at PPL Therapeutics, which contributed to the Dolly paper.

Egg-free approaches may also enable what many see as the most promising potential application of cloning: the creation of human embryonic stem cells, or cells made from them, that could be used to treat human disease.

Many scientists see human embryonic stem cells as having the most potential for the next decade, especially if scientists learn to reprogramme cells without using eggs.

"I think in ten years, people will learn how to convert one cell type to another, without nuclear transfer, without eggs, in a Petri dish," says Jaenisch. "Once we learn all the rules, maybe one day it will be possible."

Erika Check

"I think we should be surprised that cloning works at all."

2003

Italian scientists at the Laboratory of Reproductive Technology in Cremona announce Prometea, the first horse (*Equus caballus*) clone created from a skin cell, raising hopes that clones could one day perpetuate the genetic line of castrated geldings.



2003

French and Chinese scientists unveil Ralph the cloned laboratory rat (*Rattus norvegicus*). Rats had been tough to clone because rat eggs divide before the point at which the donor DNA is injected, so the technique relied on using drugs to inhibit division.



2004

Although Seoul National University researcher Woo Suk Hwang's claim to have derived stem-cell lines from cloned human embryos was later discredited, his group can still boast the most experience, and probably the highest number of cloned human embryos, but there is no hard evidence for this.



2005

Hwang's lab announces Snuppy the cloned dog. Although much of the stem-cell research from this lab has been discredited, Snuppy's clonal credentials have been confirmed.



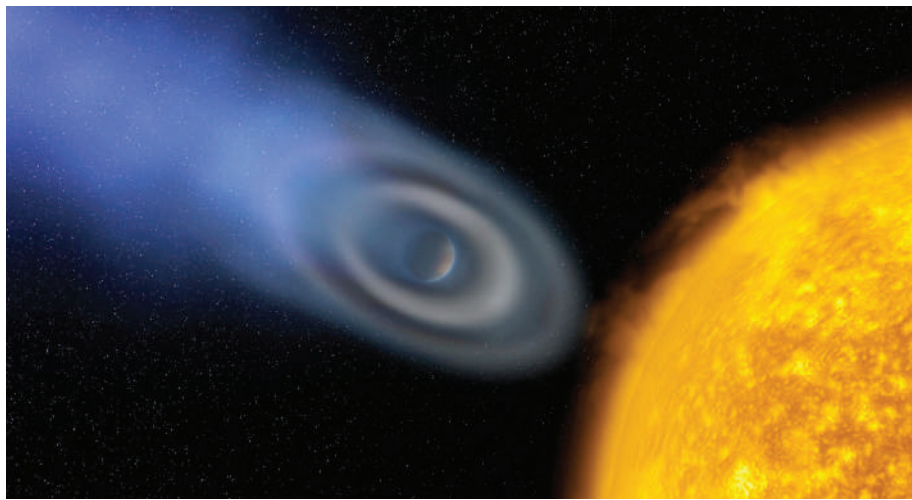
Heidi Ledford

**AAAS MEETING**

Catch up on the gossip from the conference in San Francisco.

<http://blogs.nature.com/news>

Direct view of a dark and distant world



The contents of exoplanet HD 209458b's atmosphere remain controversial.

Three groups of astronomers have made the first detailed measurements of radiation emitted by two exoplanets. Although sketchy, the results hint that the worlds are cloaked in black silicate clouds that don't let light in or out.

After a bitter race, Jeremy Richardson at NASA's Goddard Space Flight Center in Greenbelt, Maryland, and his colleagues are the first to publish their results (see 'Crossing the line'). They used the Spitzer Space Telescope to analyse the atmosphere of a hot Jupiter-like planet called HD 209458b, 150 light years from Earth, and announce their results in this week's *Nature* (see page 892).

Previous studies used the Hubble Space Telescope to look for wavelengths of starlight that are absorbed when HD 209458b passes in front of its parent star, and have deduced the presence of sodium, as well as possibly carbon and oxygen, in its atmosphere (D. Charbonneau *et al. Astrophys. J.* **568**, 377–384; 2002; A. Vidal-Madjar *et al. Astrophys. J.* **604**, L69–L72; 2004).

But Richardson's group used the Spitzer Space Telescope to detect the planet's atmosphere directly, by comparing the infrared radiation from the star and planet together, with that emitted when the planet was behind the star.

Molecules absorb and emit infrared radiation at specific wavelengths when they vibrate, and the researchers had expected the planet's infrared spectrum to show up water and carbon dioxide, as well as carbon-containing molecules such as methane. But Richardson says that the spectrum was flat, showing nothing apart from a peak he attributes to silicates, and possibly a molecule containing carbon-carbon bonds such as those seen in benzene.

The absence of water was particularly surprising because gas giants such as HD 209458b are predicted to produce large quantities of water. "Theorists will tell you that water has to be there," Richardson says. He thinks that the silicates might provide an explanation. Previous observations of the planet have suggested the presence of high clouds, so it's possible that the silicate clouds are blocking the water, and other chemicals, from Spitzer's gaze.

Carl Grillmair at the California Institute of Technology in Pasadena and his colleagues used Spitzer to measure the atmospheric spectrum of a similar planet 60 light years

from Earth, called HD 189733b. They weren't far behind Richardson in their analysis, and their full results will soon be unveiled in the *Astrophysical Journal*.

Grillmair's group also got a flat spectrum, with no water, methane or carbon dioxide. The team didn't see silicates either, but co-author David Charbonneau from the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, says that Richardson's result offers a tantalizing clue to what might be going on. "We find it hard to believe that there isn't water because it is so easy to make," he says.

Mark Swain at the Jet Propulsion Laboratory, Pasadena, California, offers a third perspective. He started working on Richardson's data once they were made public a year after they were collected, and has submitted his results to the *Astrophysical Journal*. For now he is keeping the details to himself, but he says that he didn't see any water or silicates and isn't convinced by Richardson's idea of silicate clouds. "That's not a conclusion I would sign up to," he says.

All three groups have applied to make follow-up observations with Spitzer. Theorist Alan Boss of the Carnegie Institute in Washington DC says that it's not surprising that the researchers didn't see what they expected first time round. In exoplanet research, with so few data to go on, theory must follow observations as they are made, he points out. "It would have been a miracle if theorists had been able to predict what the spectrum of these two hot Jupiters would be like."

Katharine Sanderson

Crossing the line

The debate over the presence of silicates isn't the only controversy the work has triggered. Astronomers are also discussing the race that led up to this week's publication on HD 209458b.

Jeremy Richardson and his colleagues from NASA's Goddard Space Flight Center in Greenbelt, Maryland, used the Spitzer telescope to make measurements of the planet in July 2005. By Spitzer rules, they had sole use of the data for 12 months, at which point the data were made publicly available.

Mark Swain from the Jet Propulsion Laboratory in

Pasadena, California, then began his own analysis of the data. But even though he started a year behind, he was quicker, and Richardson says that he had to rush to publish first. "It was certainly a time to redouble my efforts," he says.

Richardson isn't happy about the attempt to scoop him, especially for something as high profile as being first to measure an exoplanet's atmosphere. If you simply want to contribute to a fast-moving community, then using someone else's data makes sense, he says. But he warns that it might not be the right approach if the

goal is to establish working collaborations with members of the scientific community.

Swain says that he is surprised that Richardson's analysis took so long. "I expected they would write the paper before me," he says.

The race has raised a few eyebrows, but most astronomers seem to agree the most important thing is to get the results out fast. "It isn't bad form," says Geoffrey Marcy, from the University of California, Berkeley, who discovered 121 of the 209 exoplanets found so far. "Spitzer is expensive and we want the best results." **K.S.**

ESA/A. VIDAL-MADJAR, PARIS ASTROPHYS. INST.

Sticky situation: a programme falsifying the health benefits of *natto* has caused an outcry in Japan.



DIGITAL ARCHIVE JAPAN/ALAMY

Japanese TV show admits faking science

TOKYO

For scientists who speak to the mass media about their research, a scandal surrounding a Japanese television show demonstrates just how badly things can go wrong. Kansai Telecasting Corporation (KTV) has admitted that it faked scientific results in one of its programmes, and used dubbing to put false words into the mouths of foreign researchers. After an outcry from the Japanese media, the company has cancelled the series and is reviewing all 520 episodes.

One of the victims is Dr Kim, a professor at a US university who studies soya bean fermentation. He spoke to *Nature* but asked that his first name and affiliation be withheld. He was interviewed by KTV about his research in January 2006. When he finally watched that programme this month, he was horrified. The show — and Kim — trumpeted the health benefits of a fermented bean paste called *miso*. “They made it look like I said many things that I didn’t. About 60% of the content was not correct,” he says.

Hakkutsu: Aruaru Daijiten II (which roughly translates as “Excavation: an encyclopedia of facts”) was a popular weekly show broadcast for ten years by the Fuji Television Network

and watched in around 15% of Japanese homes. Presented as a science documentary, the show covered food and health-related topics and was based on interviews with scientists.

The scandal broke after the broadcast on 7 January, which claimed that eating *natto*, a pungent dish made from fermented soya beans, can help you lose weight. When the Japanese media raised questions, KTV investigated, and on 20 January released a statement admitting that most of the data in the programme had been fabricated. Photos of slimmed-down people were not of the experimental subjects, and the claimed measurements of reduced fat levels had never been made.

KTV admitted that it had used Japanese dubbing to put words into the mouth of US researcher Arthur Schwartz of Temple University in Philadelphia. In the 7 January programme, Schwartz appeared to say that isoflavone, a compound in soya beans, can increase the body’s levels of dehydroepiandrosterone (DHEA), a naturally occurring hormone said to have various health benefits. Experiments portrayed as having been carried out by Schwartz were

actually those of another researcher.

Schwartz, who is studying the ability of DHEA to slow the progression of age-related diseases, did not respond to *Nature*’s request for an interview, but a university spokeswoman confirmed that “statements were wrongly attributed to Dr Schwartz and his research was falsely represented”.

Other incidents have since come to light, including the episode of *Aruaru* that changed Kim’s words. “It looks like I am saying that *miso* can make you thin and even that it has certain neurological effects,” Kim says. He had actually been explaining why fermenting soya beans makes them more easily digestible and useful as animal feed. “I spent a half-day taking them round my laboratory and showing them results. They understood what I was doing,” he says.

Scientific impropriety has been traced back to the early days of the show. In 1998, *Aruaru* asked Yoichi Nagamura of the Chiba Institute of Science, who researches the soporific effects of tryptophan, to run an experiment investigating whether lettuce makes mice sleepy. Nagamura and the KTV production crew fed

“They made it look like I said many things I didn’t.”

lettuce extract to some mice and watched them for two hours. "I was hoping to be on TV, and I hoped the mice would fall asleep," he says. But he and the crew agreed that there was no noticeable effect.

So Nagamura was surprised when he saw the programme show one of his mice, declaring: "It's fallen asleep!" Makoto Tajima, a nutrition researcher at Jissen Women's University in Tokyo, then appeared explaining that lactucopicrin, a chemical found in wild lettuce, and in trace amounts in cultivated lettuce, can induce sleep. "The programme left the impression that eating three leaves of lettuce can knock you out," says Nagamura.

Tajima says he's never carried out any experiments with lettuce, but that he gave accurate information from the scientific literature. He says he felt "a little uncomfortable" explaining another scientist's results, but wasn't too concerned: "We're used like TV personalities, I say what the programme wants me to." Tajima adds that he has appeared on more than 500 television programmes to explain nutrition. "If I didn't do it, they'd get someone worse."

Nagamura says he didn't complain to the production or broadcasting companies, because "it was too ridiculous". But he expressed his misgivings about the show at academic conferences and public talks. "Many scientists would say that *Aruaru* is dodgy, but nothing happened until now," he says.

Newspaper editorials in Japan are questioning a TV culture in which 'infotainment' is increasingly popular, and contract production companies are chasing ratings, particularly for shows on health and beauty.

Aruaru was cancelled in January because of the scandal, and KTV has asked an independent committee to review all its episodes. It is due to report in mid-March. A 2004 show about the 'science of face-thinning' and a 2001 show about the power of adzuki beans to invigorate the brain are reportedly getting close attention.

Tomohiro Takeuchi, assistant chief of the terrestrial broadcasting division of Japan's internal affairs and communications ministry, says that KTV has technically broken the country's communications law, which prohibits "perversion of the truth", and that his division is waiting for an internal report from KTV, due by the end of February, to decide what action to take. But the ministry's only options are to revoke KTV's broadcasting licence, which Takeuchi says is generally thought too harsh, or censure

by "administrative directive/guidance". The stricter penalty has never been applied, and the lesser one in only a few cases. Last year another production company was censured after one of its programmes on how to parch white navy beans for most effective weight loss led to viewers being poisoned.

Earlier this month, the minister of internal affairs and communications called for the next parliament to add other measures, stricter than a directive but less harsh than revoking the licence, to prevent bogus reporting.

Kuniko Takahashi, a nutrition researcher at Gunma University in Maebashi City, says the problem is widespread in Japan. Since 2000 she has been following programmes that combine health and nutrition with entertainment, and has found exaggeration and some dangerous mistakes. "They should all be treated with suspicion," she says. Nevertheless, she worries about the proposed law change. "It would let the government tell us what was right and wrong," she says. "In the end, we'll have to depend on the morals of the people in the industry."

On 26 January, Ichiro Kanazawa, a neurologist and president of the Science Council of Japan, issued a statement saying that the council's guidelines on academic fraud should apply to anyone doing experiments, even for television shows. He's not optimistic the guidelines will be followed by TV companies, however. "Their goal is not scientific truth; it's ratings," he says. "They are under



Tall tale: the mouse that didn't fall asleep.

tremendous pressure."

Cases as blatant as *Aruaru* are rare, but the temptation to wow an audience at the expense of scientific integrity is not confined to Japan. Last year, *Brainiac: Science Abuse*, a popular UK science series, admitted that in an episode claiming to show a bathtub of water being blown up by caesium and rubidium, explosives were used instead. Sky One, the channel that broadcast the series, released a statement saying that although it consulted with scientists when planning stunts, its priority in executing them was entertainment. "If you're expecting to see the Open University, you're on the wrong channel!" it said.

Both Kim and Nagamura are now wary of any media. When consenting to be interviewed by *Nature*, Kim insisted he saw how he was going to be quoted: "That is what I learned from the *Aruaru* experience."

David Cyranoski

NUMBER CRUNCH


10,400 metres is how far Google Maps Australia recommends you should travel to get from 200 Sussex Street in Sydney across the road to 201 Sussex Street.

Aus\$3.00 is the toll you must pay (roughly US\$2.35) to cross Sydney Harbour Bridge as part of Google's suggested route.

300 metres is the actual, toll-free, distance between the two buildings — one of which just happens to be Google's Sydney headquarters.

SCORECARD

 **Chinese romance**
Shanghai restaurateurs are serving sweet potatoes grown from seeds that went into space on last year's Shenzhou VI mission. The star-crossed spuds are purple — the traditional Chinese colour of romance.

 **Nagging**
Constant nagging really does seem to be counter-productive. In a recent experiment, volunteers completed anagrams more slowly when the face of a nagging loved one was flashed subliminally on a screen. Faces of fun-loving friends and family had no such effect.

ON THE RECORD

“Danger, stay away.”

The message conveyed by the new ‘crystal clear’ radiation warning sign designed by the International Atomic Energy Agency. Looks good, although it does

seem to advocate running in the workplace.

“Tenure is not an inoculation against foolishness.”

University of Maryland physicist Robert Park's opinion on the work of Robert Jahn, whose Princeton University tenure allowed him to run a lab investigating telekinesis.

Sources: TechCrunch, BBC, J. Exp. Soc. Psychol., Reuters

Role of state climatologist comes under scrutiny

Many climate scientists get frustrated with those who don't believe that human activity is causing global warming, but should having such views be a sackable offence? In recent months, two US state climatologists have been asked to stand down from their posts because of it, triggering debate about whether personal views should determine suitability for what many see as an academic position.

Almost all US states have their own climatologist, who runs a small office concerned with meteorological matters of local interest. Earlier this month, Oregon governor Ted Kulongoski said that he wants to strip Oregon's climatologist George Taylor of his title for not agreeing that global warming is predominantly caused by humans.

Taylor, based at Oregon State University in Corvallis and appointed to his post in 1991, argues that his post is academic rather than political, and that it's not his job to tell the state government what to do. “Most state climatologists have never even met their governor, let alone offered policy advice,” he told *Nature*. Therefore, he says, his personal views on climate change shouldn't be an issue.

But Kulongoski clearly believes that a state climatologist should represent the state, and he argues that Taylor's views are inconsistent with Oregon's goal of cutting greenhouse-gas emissions. “He is Oregon State University's climatologist,” Kulongoski said in a television interview on 8 February. “He is not the state of Oregon's climatologist.”

Patrick Michaels, Virginia's climatologist since 1980, faces a similar situation. An environmental scientist at the University of Virginia, he does not agree that taking action to cut greenhouse-gas emissions is an urgent priority. Last August, he received a letter from the office of the governor, Timothy Kaine, asking him not to use the title any more.

Most state climate offices, like those of Michaels and Taylor, are based at state universities, although a few are directly within state government. Thirty-four of the state climatological offices are officially recognized by the American Association of State Climatologists (AASC), which sets standards for state climate offices, but cannot bestow or remove 'state climatologist' status. To qualify for AASC

recognition, an office must make its data public, set up educational initiatives and work with the media. It must also monitor climate conditions, place them in historical context and assess possible future impacts on the climate.

The role is somewhere between that of a meteorologist and a climate modeller, and state climatologists are concerned only with their own state. Residents of New Jersey, for example, can thank their state climatologist David Robinson for the knowledge that they have just encountered their warmest December since records began in 1895.

So grand ideas about the need to tackle global warming aren't the main focus, says Mark Shafer, a climate officer who works with Oklahoma state climatologist and former AASC president Ken Crawford. Although stopping short of saying that state climatologists shouldn't be sacked for their views on global warming, he points out that they generally deal with questions to which those views aren't relevant — for example, the impact of human actions on shorter-term climate effects such as drought.

The AASC is now preparing a statement in response to the controversy, urging state governors to consider the factors most important for a successful state climate office. “We encourage public officials to enter into a direct dialogue with their respective state climatologists in order to resolve any differences of opinion,” says a draft of the letter, seen by *Nature*. Shafer says there are roughly a dozen prospective signatories so far.

Many state climatologists are unhappy about the growing importance of global climate issues on their agenda, however, and the way state climatologists' views on global warming (particularly those of the sceptics) are increasingly sought by the media. Taylor points out that the role initially grew from the National Weather Service in the 1950s, and has always been more closely akin to a weather information service. “Climatologists must be knowledgeable about climate change, but that is usually only a very small part of what we do,” he argues.

All of which moves the debate towards a different question. As the priorities of a state change, should those of its climatologist?

Michael Hopkin

“Most state climatologists have never even met their governor, let alone offered policy advice.”

China charges drugs lab with making illegal Tamiflu

Six people are on trial in Shanghai for making counterfeit versions of Tamiflu, the drug thought to be the most effective treatment for avian influenza.

Xun Wang, head of a pharmaceutical operation known as Shanghai Xidi, is accused of running a lab that produced Tamiflu 24 hours a day between November 2005 and May 2006. Official Chinese news sources said on 9 February that Wang had obtained instructions about how to produce the drug. It is not clear whether the counterfeit pills contained the correct ingredients. The authorities were alerted to the alleged fraud via Internet sites that advertised the drug. More than 400 kilograms of pills were subsequently seized.

Countries around the world have been stockpiling Tamiflu to prepare for a possible pandemic. China's health ministry seized control of distribution of the drug last year amid fears of hoarding and piracy.

Puzzle-solving quantum computer is unveiled

A Canadian firm has revealed what it claims is the first fully functioning quantum computer — generating both interest and scepticism from physicists.

D-Wave Systems, based in Burnaby, British Columbia, debuted its system on 13 February at the Computer History Museum in Mountain View, California. The computer used its 16 quantum bits, or qubits, to match proteins in a database, create a seating chart for a wedding party and solve a sudoku puzzle.

Critics say that the machine, which takes an unusual approach known as

'adiabatic quantum computing', may not be performing strictly quantum-mechanical computations. The adiabatic technique leaves the machine to conduct quantum computations on its own, making it difficult to tell whether it is behaving in a quantum or a classical manner.

"I'm really very sceptical," says Umesh Vazirani, a computer scientist at the University of California, Berkeley, adding that he would like to see more data before he is convinced.

Kansas puts evolution back on the curriculum

The Kansas State Board of Education has scrapped standards that many scientists said supported intelligent design, a concept that holds that a supernatural power shaped evolution (see *Nature* 436, 899; 2005).

On 13 February, the board decided by six votes to four to replace the current guidelines with a set that strongly supports the theory of evolution and that was written and endorsed by scientists and educators. Several board members who supported the old standards lost their seats in primary elections last August, which all but ensured this month's reversal.

The decision is a victory for supporters of evolution, says Jack Krebs, who heads Kansas Citizens for Science, an advocacy group that fought for the changes. "Kansas has once again returned to the fold of mainstream science," he says.

Biologist calls time on hunger strike at MIT

An African-American stem-cell researcher has ended a hunger strike that was designed to protest against racism at the



Food for thought: James Sherley has highlighted the issue of race in academic appointments.

Massachusetts Institute of Technology (MIT) in Cambridge.

James Sherley held his strike outside the provost's office for 12 days before accepting solid food on 16 February. Sherley, who began the action after being denied tenure at MIT, said he was ending "in celebration of the attention that has been brought to bear on issues of equity, diversity and justice at MIT and in higher education".

A statement from MIT acknowledged that Sherley's protest had "focused attention on the effects that race may play in the hiring, advancement and experience of under-represented minority faculty". MIT president Susan Hockfield has already promised to establish a committee to study potential racial bias in the institute's hiring and career-advancement procedures. But officials said that no change had been made regarding the decision on Sherley's tenure, which they maintain was made fairly.

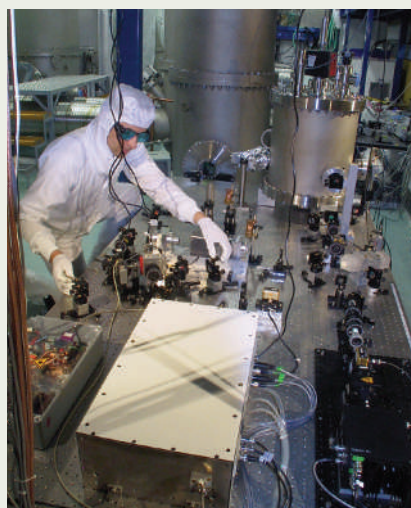
Healthcare company takes lead with biobank

A private healthcare organization in the United States has announced plans for one of the largest studies of how genes and lifestyle affect health.

Medical researchers are keen to collect data on large cohorts of people to help them tease apart the genetic and environmental factors that cause some people to develop diseases such as cancer. Kaiser Permanente, a non-profit healthcare provider based in Oakland, California, announced on 14 February that it is inviting about 2 million of its members in northern California, who come from diverse ethnic backgrounds, to add data to the new bank. Participants will submit a DNA sample, medical history and lifestyle information.

A huge cohort study proposed by the US National Institutes of Health (NIH) is still some way from recruiting participants, although Kaiser Permanente says that it will collaborate with NIH researchers on the project.

Labs join forces to detect gravitational waves



Physicists have decided to pool their data in an effort to detect gravitational waves, the perturbations of space-time that are predicted to ripple away from events such as the merger of black holes.

The teams behind the three biggest projects to search for such waves announced the move on 13 February. If all the facilities detect a signal, it would increase the teams' confidence that their findings are valid. In addition, slight differences in the time that a signal arrives at the different sites should help to pinpoint the source.

The agreement was struck between the Virgo detector, based in Italy, the three Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors in the United States, and the GEO600 detector (pictured) in Germany. Virgo aims to begin its first science run in May; LIGO and GEO600 are already gathering and sharing data.

BUSINESS

Spy in the sky

Except for their use in military reconnaissance, unmanned aircraft have been seen mainly as expensive toys. But as technologies mature, **Ned Stafford** asks when drones are set to go commercial.

Picture this: in the middle of the night, you hear a prowler outside your home and call the police. In less than a minute, even as a patrol car is dispatched, an aerial mini-drone equipped with night-vision cameras has arrived overhead to scout the scene, sending back images to the police as they approach.

If this sounds like science fiction, it won't for much longer, says Patrick Egan, membership director of the Remote Control Aerial Photography Association, based in Scottsdale, Arizona. "Within 10 years, this will be common practice," he claims.

Drone aircraft, also referred to as unmanned aerial vehicles or UAVs, have gained prominence of late because of their extensive use by US military and spy agencies in Afghanistan and Iraq. But according to the scientists, engineers and enthusiasts who have been playing around with them for decades, that military use reflects real technical improvements that could see drones acquire all sorts of other uses as well.

"The current state of UAV technology is comparable to the early, barnstorming days of manned aviation," says Kee-Choon Wong, an aeronautical engineer and leader of a UAV research group at the University of Sydney, Australia. "Military applications of UAVs are beginning to mature, and some of the technologies will flow into commercial applications. We have only seen the tip of the iceberg of what is achievable."

Shaping up

Today's non-military UAVs come in all shapes and sizes, weighing from as little as a few hundred grams to more than 100 kilograms. They can have fixed or rotary wings, and are powered by either fuel engines or electric cells.

Electric-powered drones are generally quieter and more reliable than fuel-powered ones, but have a shorter range. UAVs with fixed wings are easier to build and generally cost less than the helicopter version. They also have a greater range and can reach faster speeds. The bigger they are, the higher, longer and further they can fly and the more they can carry, but they also require more ground support. Smaller systems are portable, can be launched into the air by hand, and are easily recovered.

"The non-military market is insignificant."
— Shai Shammai



Unmanned helicopters can be fitted with cameras to send images to crews on the ground.

But regulatory uncertainty has made it tough for drones to obtain a real foothold commercially. Drones used by hobbyists are generally banned from commercial airspace — meaning, in effect, that the vehicles cannot fly higher than a kite, or drift out of eye contact of the person operating them.

Serious commercial applications will need permission to fly further than that. So the International Civil Aviation Organization has established an informal group to discuss the regulation of commercial drones, according to spokesman Denis Chagnon. He adds that the next step will be to establish terms of reference for a formal group, which could meet as soon as this September. "There is a need to come to a global consensus on the use of UAVs," says Chagnon.

In the meantime, outside observers say that the market for commercial drones remains undeveloped. Shai Shammai, an analyst in the London office of the market-research firm Frost & Sullivan, says: "The non-military market is insignificant." A study in 2005 by Frost & Sullivan for the European Union estimated that the European market for UAVs would be worth just US\$1 billion over the next ten years. But "some would say that's an optimistic view," notes Shammai.

Daryl Davidson, director of the Association

for Unmanned Vehicle Systems International in Arlington, Virginia, admits that the non-military drone market is small and ill-defined. "But that doesn't mean it won't happen," he says, adding that in his view, pessimism about the market is driven by 'near-sightedness,' as well as by uncertainty about regulation.

Captive markets

If commercial drones do take off, four groups of businesses would be looking to cash in. Academic researchers, such as Wong, have associations with small, specialist companies that build UAVs. Older commercial companies, often set up by hobbyists, have long sold drones as toys. A handful of major corporations already have a toe-hold in the market. And military contractors have perfected the secret designs of the world's best-performing drones — those already used by air forces and spy agencies.

Civilian applications have been a long time coming. Drones have been around for almost as long as manned aircraft. The first drones, developed during World Wars I and II, tended to be the size of aeroplanes; smaller ones were developed for the US military during the Vietnam War.

Miniature aircraft are tricky to fly and control, however, partly because of the aerodynamic properties of their wings and rotors. "The air encountered by a Boeing 747 just gets

MICRODRONE

out of the way, whereas the same air encountered by a small insect is like swimming through honey," explains Robert Michelson, an engineer who runs the consultancy firm Millennial Vision, in Canton, Georgia.

To improve their designs, drone manufacturers have been taking advantage of new technologies. For example, ever smaller and lighter cameras can be incorporated into the craft to beam real-time images back to the ground. Sensors to measure radioactivity or other environmental conditions can also be incorporated cheaply if required.

Drones have also become much easier to fly: better gyroscopes and control software help to stabilize them during flight, and the global positioning system simplifies remote navigation.

Four-rotor, remote-control helicopters made by Microdrones of Kreuztal, Germany, for example, use three gyroscopes and three accelerometers each for stabilization. Their MD4-200 craft weighs less than a kilogram and can fly for 20 minutes between charges.

Microdrones was set up in 2005 by four men who liked to fly remote-control hobby aircraft. The company now has 12 employees and has sold more than 100 of its helicopters for about €10,000 each (\$13,000).

Virtual reality

Thorsten Kanand, one of Microdrone's founders, says that the MD4-200 is much simpler to fly than drones modelled on traditional, single-rotor helicopters. The person handling the remote control can even wear goggles to see live images relayed from the vehicle hovering overhead. "This is the sort of gadget we saw in science fiction shows on TV just a few years ago," he says. "But it is reality now."

Kanand notes that many buyers are as enticed by the fun of playing with the drones as by their practical value. "It's not always the businessman I am selling to," he says. "It is often the small boy in the businessman."

Drones have already established a modest foothold in some areas. In Japan, for instance, Yamaha Motor sells up to 300 of its RMAX unmanned helicopters every year for crop-spraying on private land.

But enthusiasts foresee a bolder future. Michelson, for example, predicts that major improvements in propulsion and autonomy will create "thinking machines" of "sizes down to a few centimetres, with the endurance of a flying insect." And where Michelson sees small, Davidson thinks big. "When the commercial segment really takes off, you'll see unmanned aircraft begin to take on almost any traditional role that is done by a piloted aircraft," he says. ■

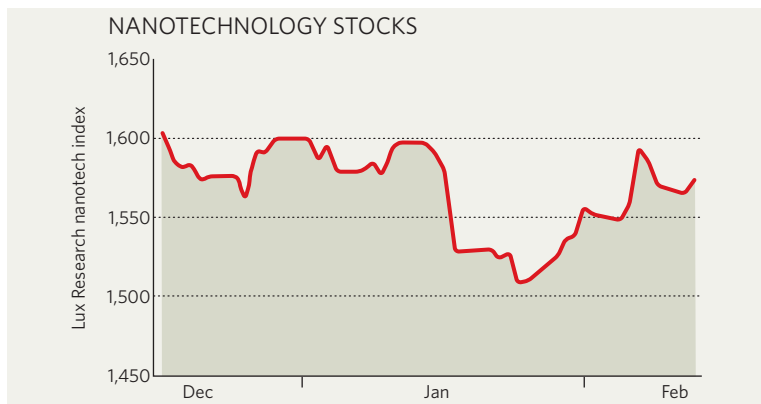
IN BRIEF

CLEAN FUEL SHIFT Diversa, the San Diego bioprospecting company, is acquiring a biofuels specialist — Celunol of Cambridge, Massachusetts — and shifting its corporate focus to clean energy. Diversa says that it will continue to try and develop drugs based on natural compounds, but will move its headquarters to Cambridge. It intends to retain an undetermined number of its 200 personnel in San Diego to do research and development. The \$180 million acquisition of the privately held company is expected to be completed in June.

DRUG DROUGHT The US Food and Drug Administration (FDA) approved just 18 new drugs — ones whose active ingredient has not been marketed before — during 2006. The number of new approvals is the same as that for 2005, and the worst in recent history, apart from 2002, when only 17 new compounds were approved. In 1996, 53 such drugs won approval. Since 19 prescription medications came off patent in 2006 and became available as generics, the industry had fewer patented products on the market at the end of the year than at the start. The numbers were compiled by RPM Report, a monthly drug regulation and policy newsletter.

ALTERNATIVE CLIMATE The United States has extended its lead as the most favourable location for 'clean' energy businesses to make money, says an annual 'attractiveness survey' by Ernst & Young. The London accountancy firm says that President Bush's support for biofuels and a raft of subsidies and other measures aimed at cutting oil imports, together with the sheer size of the market, makes the United States the most promising location for such companies. India and Spain came second-equal in the rankings.

MARKET WATCH



The new year has seen some auspicious nanotechnology announcements from major corporations. On 6 February, Eastman Kodak, based in New York, rolled out nanotechnology-based ink and jets for printers. Hewlett-Packard, IBM and Intel have all announced progress in microchip performance, based on innovation at the nanoscale.

Not all this excitement, however, has been reflected in the Lux Nanotechnology Index, which tracks the performance of companies that produce and use nanotechnology products.

The index started the year drifting down, but picked up a little in February. Peter Hebert of Lux Research, the New York consultancy that compiles the index, says that good and bad company news has been pulling the index in opposite directions.

On the plus side, Oregon-based FEI,

which makes scanning electron and focused ion-beam microscopes, saw its shares leap ahead from US\$25 to \$33 after results were released on 7 February. Its sales for the last quarter of 2006 were up 41% from the year before, to \$140 million.

Investors were less impressed by NVE Corporation, a small, Minneapolis-based specialist in spintronics, and research-tools supplier Symyx Technologies, based in California. Although both companies announced that investors' profits and sales were up, the rises were not enough to satisfy their investors.

Hebert sees these companies' results as boding well. FEI's results in particular, he says, show that corporate demand for nanotechnology tools "is really ramping up".

Colin Macilwain

GHOST BUSTER

An Italian scientist revived the hunt for the mafia's boss of bosses. **Declan Butler** reports.

Bernardo Provenzano, Italy's most wanted man, was on the run for more than 40 years. Moving from one safe house to another, he became the godfather of the Sicilian Mafia in 1993, running the huge organization without ever using a telephone, communicating instead through tiny typed notes, *pizzini*, passed back and forth through a labyrinth of secret relays.

Over time, many outside the Mafia had begun to believe the rumours, and his lawyers' assertions, that he was dead. That was until an Italian scientist used DNA to confirm that Provenzano was still alive, relaunching the manhunt that led to his capture on 11 April last year.

Provenzano had risen up the ranks of the Mafia by mowing people down, earning him the nickname 'the tractor'. He was a member of the Corleone clan, which took control of the Sicilian Mafia after a power struggle in the early 1980s that left hundreds dead. Provenzano was the right-hand man, and natural heir, of Salvatore Riina, who was the Corleone boss and Sicilian godfather until his arrest in 1993.

The invisible enemy

The prosecutor who was most responsible for Riina's arrest, Giovanni Falcone, was murdered in 1992, and Provenzano was implicated in the crime. Pietro Grasso, who is now Italy's national prosecutor against organized crime, vowed to find Falcone's killers. But, as Grasso acknowledges, all the police had to go on was a photograph of Provenzano taken in 1959, and a computer-generated 'identikit' of what he might look like now — his invisibility led to another nickname, the ghost of Corleone. With few leads, there followed a decade of fruitless searching.

But Grasso never gave up. In 2005, he asked Giuseppe Novelli, head of the medical genetics lab at Tor Vergata University in Rome, to help identify Provenzano. "I was taken aback at first," says Novelli, a specialist in genetic fingerprinting. "I mean, how can you identify someone who you have no information about?"

But Grasso had a lead from a Mafia informant, or *pentito*, arrested in 2005. In 2002,



Wanted: Bernardo Provenzano in 1959; in 'identikit'; and on his capture last April.

the source said, Provenzano had gone to Marseille in France under a false name to be treated for a prostate tumour. A police raid at the clinic found a case history for a man named Gaspare Troia, who had never been admitted to the hospital. The patient's description matched Provenzano.

Novelli extracted DNA from hospital tissue samples and ran a DNA profile of both mitochondrial and Y-chromosome DNA. He then compared the DNA with that taken from blood samples of Provenzano's brother, stored in a Palermo hospital where he had been admitted for surgery.

The results came in November that year. "It showed that they shared the same mother and father — they were brothers," says Novelli. "The lead was the right one, they knew they had the right man."

"Novelli's group did a terrific job," says Grasso. "You cannot even imagine my enthusiasm when I had the proof of the complete match between the profiles." The result gave prosecutors and police the strength to carry on looking for Provenzano, adds Paola Di Simone, head of the Italian Scientific Police's Forensic DNA Unit in Palermo.

Knowing that Provenzano was alive, the investigators focused their attention on busting the ring of people who had provided the logistics of Provenzano's trip and hideout in Marseille. Last April, by tracking a packet of laundry sent by his wife, police captured Provenzano in an isolated farmhouse just a mile from his birthplace in Corleone.



L. BRUNO/AP

AP

At the hideout, Di Simone collected DNA from false teeth, glasses and an electric razor to confirm Provenzano's identity. She also hoped to find DNA in the house pointing to the henchmen, but the only DNA was his. "Being in the den just after his capture was very emotional," she says.

Code of silence

Novelli told no one what he was working on, except his lab collaborator Ruggiero Mango, who also worked on the case. Grasso advised secrecy to protect the scientists and their families. Novelli admits to having worried for his safety. "I spent one year in a dangerous situation," he says, recalling flying once to Palermo to meet Grasso and being whisked away from the steps of the aircraft in a car.

"I didn't even tell my wife," says Novelli. "She learnt it from the TV after Provenzano had been captured." But he now speaks openly about the work and says that he is no longer concerned for his safety. "It's finished, he is in prison. I think it is over — I hope so."

Provenzano was convicted *in absentia* in 1993 to life imprisonment for his involvement in the bombings that killed Falcone and another prosecutor, Paolo Borsellino, in 1992. He now faces further trials on other charges of murder and blackmail.

His capture also marked the end of an obsession for Grasso, who says he is overjoyed "to have fulfilled the promise I made on the grave of Falcone, that for the rest of my life I would pursue a sole objective: capturing all those responsible for his death."

Declan Butler is a senior reporter for *Nature* based in Paris.

"I spent one year in a dangerous situation. I didn't even tell my wife."
— Giuseppe Novelli



MUDDY WATERS

How did a mud volcano come to destroy an Indonesian town?

David Cyranoski reports from Sidoarjo.

It started on 29 May 2006: a small spurt of mud in the middle of a rice paddy. Now the cauldron of hot, bubbling mud is some 50 metres in diameter and rises 16 metres above that long, submerged paddy. A dented horseshoe of a levee — the tallest of a series of often-failed efforts to protect homes and factories — tries to channel the muddy outflow towards a river. On the disaster-management maps that show those 13 kilometres of levee and 450 hectares of muddy devastation, the source at the centre is labelled simply 'Big Hole'.

The Big Hole, more commonly called Lusi — a contraction of *lumpur* for mud and Sidoarjo, the place in East Java where the mud is erupting, is the opening of what's called a mud volcano. It's not, perhaps, the most apt of names: the mud is more like thick water, and the volcano's only structure is that imposed on it by the artificial embankments. But if it gets across the idea of something bizarre and disastrous, then the name is doing its job. In the Big Hole, below the thick,


billowing plumes of vapour, black waves ripple the unfathomable surface. "It's like the sea," says my guide — but a sea that is pockmarked with bubbles of gas from a source about which little is known, even after nine months.

A smooth black lake stretches for more than a kilometre in most directions, past treetops, past street lights, and over barely exposed roofs. The flow, which started at just 5,000 cubic metres a day, has now topped 130,000, and Lusi has already displaced 24,000 people. To the southwest, within an area closed off by levees, the mud has crusted over into chunky, grey blocks that come up to the eaves of a series of factories. In an earthwork-protected area that allows access to the Big Hole, bustling workers dismantle a drilling rig that tried and failed to intercept the mud surge on its way to the surface. On 22 November, the weight of the mud destroyed a gas pipeline, killing 13 disaster workers; two other workers have since died in accidents with heavy equipment.

And no one knows when the mud flood will stop: Richard Davies, a geologist at Durham University in the United Kingdom with an interest in mud volcanoes, points to some that have been spewing forth for months or even, as in the case of the Koturdag mud volcano in Azerbaijan, decades. "I expect Lusi to be bubbling for years to come," he says.

Lusi's devastation has so far been a defeat not only for civil engineering but also for scientific understanding. Efforts to figure out how it started have turned into a hotly contested whodunit with two suspects: a drilling project by an Indonesian oil-drilling company named PT Lapindo Brantas, and a nearby earthquake. Now Earth scientists are starting to weigh in. Geologist Adriano Mazzini at the University of Oslo in Norway sees it as a golden opportunity to learn about the evolution of these strange phenomena, which are generally studied only long after the structure has set. "We can see this one from day one," says Mazzini. But such

D. ARDIAN/GETTY IMAGES



Production line: workers battle to save what they can from what used to be their homes and workplaces.



studies are hindered by the fact that Lapindo has been keeping much of the drilling data under wraps. Meanwhile, the national team charged with sorting things out has focused its efforts on penning in the mud. It has had little time or effort to spare for obtaining the data on temperature, viscosity and flow rate that scientists might be interested in. “We are trying to save homes. This is not a science experiment,” says Basuki Hadimuljono, the head of the team. But with no signs that the mud will let up, and the possibility that a greater understanding of what happened could improve the countermeasures, the need to learn more is growing urgent.

What's in a name?

An estimated couple of thousand mud volcanoes have been seen around the world, including several in East Java. Mostly they are a naturally occurring process in which sediment that has been buried deep in the Earth is liquefied and squirted back to the surface, along with water and gas. But little is known about the high-pressure systems that drive such outbursts and the circumstances that start them off. Some — like the Koturdag in Azerbaijan — live up

to the promise of their name, with thick mud flowing from a raised, central crater. And many are genuinely, viscously muddy. “Some of them come out a couple of centimetres per day, like toothpaste,” says Davies.

Lusi is exceptional in its sheer volume. It is also an outlier at the dilute, watery end of the volcanoes' viscosity spectrum. Its mud is about 70% water, according to Bambang Istadi, the exploration manager at PT Energi Mega Persada (EMP), Lapindo's parent company. After examining fossils within the mud, Istadi says that the particles in the mud come from a layer of shale and mudstone at a depth of somewhere between 1,220 and 1,830 metres. Davies says this mudstone continues down, mixed with sandbeds, to nearly three kilometres. Below that is the Kujung limestone formation, into which Lapindo was planning to drill in search of gas.

The first published analysis of the mud volcano, from a group led by Davies (R. J. Davies *et al.* *GSA Today* 17, 4–9; 2007), conjectures that the water driving the mud volcano comes from that Kujung limestone, and suggests that the escape could have been caused by Lapindo's drilling at a site called Banjar Panji-1. The Banjar Panji-1 well is an exploratory well, started when little was known about the underlying geology. In his paper, Davies argues that the drill at Banjar Panji-1 punctured the Kujung, allowing high-pressure water and gas to escape into the borehole. The fluids forced their way into the surrounding rock and fractured it and the high-pressure water passing through these fractures liquefied the surrounding shale before new cracks gave it access to the surface. The cracks have been growing ever since.

As evidence, Davies cites abnormally high pressure readings for the Kujung formation taken some five kilometres away from Banjar Panji-1 and for a shallower layer right at the drilling site. Davies notes that he cannot say how he acquired these data, or present more data to back up his analysis because his source must be kept confidential.

Istadi, who oversaw the drilling of the well, describes Davies' paper as “preliminary conclusions, [with] interpretation based on an incomplete data set”, and says that Davies did

not contact his company to confirm the data. Istadi's main gripe is that drilling records show that the borehole never penetrated the Kujung formation. When the team hit an instability at 2,834 metres — around the depth that they had expected to find the Kujung — they too assumed that they had reached the limestone. “At first we thought so, but it can't possibly be,” says Istadi. As evidence, he offers drilling cores that show no evidence of the 12-metre-thick layer of hardened clay that has been found draped over the Kujung elsewhere, and says that chloride levels in the water are twice as high as would be expected from the Kujung. Davies has since been contacted by Lapindo representatives who corrected a few details in his analysis, but says that nothing they told him changed his overall interpretation. He does, however, agree that the available evidence does not necessarily show that the drill entered the limestone.

Shaken, not stirred

Istadi offers another explanation for the volcano's origin. The day before the volcano erupted, a magnitude-6.3 earthquake struck Yogyakarta, 280 kilometres to the southwest of Sidoarjo. Istadi says that seven hours after the earthquake, drilling fluid, which is circulated up and down the borehole to keep the pressure higher than that of the fluids in the surrounding rock, leaked out. He thinks this drilling loss, or ‘loss of circulation’, as it is known, was caused by shock waves from the earthquake — and that the same shocks might have triggered the mud volcano: “Faults became open, lost their sealing capacity, and became permeable.” Those faults, he says, “served as the conduit where the mud flows out”. This explanation squares well with the views of one of the country's most powerful and

richest men, Aburizal Bakrie, the Coordinating Minister for People's Welfare. Bakrie, whose family owns part of Lapindo, has long been arguing that Lusi is just another “natural disaster” — no more the fault of an individual company than

the earthquake itself, or the floods that have hit Jakarta in the past months.

But others have their doubts. Davies thinks that had the eruption, or the drilling loss, been triggered by the earthquake, it would have started much more promptly. Michael Manga, a geophysicist at the University of California, Berkeley, says a causal link between the earthquake and the mud volcano would be extremely surprising. Manga has collected data on the distances over which earthquakes have had similar impacts on Earth's plumbing, triggering mud volcanoes and other ‘liquefaction events’. In 343 such events, Manga found a clear lower limit

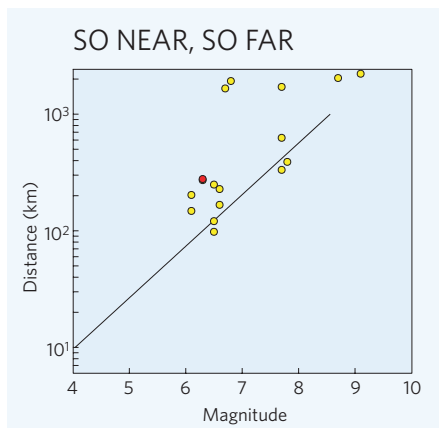
“We are trying to save homes. This is not a science experiment.”
—Basuki Hadimuljono

below which earthquakes won't do the job (see graph). "If the Yogyakarta earthquake caused that mud volcano, it would have been way out of range," says Manga. Manga also found some recent earthquakes larger than 6.3 magnitude closer to the Big Hole that did not cause a mud volcano, suggesting that such weak seismic stress could not account for Lusi.

A tragedy of errors

Moreover, the drilling loss might have started a series of events that could implicate Lapindo further, suggests Andang Bachtiar, a consultant at Jakarta-based Exploration Think Tank Indonesia. The drill team reacted with a special drilling fluid to seal the fractures responsible for the loss of fluid. The next step was to pull the drill out and add cement to the unstable area at the bottom to secure it before they continued drilling. But at 8 a.m. on 28 May, when they were pulling the drill out, they got some 'kick' 1,293 metres down. Kick is basically the opposite of loss: instead of drilling fluids leaking out of the borehole, highly pressurized liquids or gas suddenly rush in. "The monster from below caught up with them," says Bachtiar.

An executive at EMP says that kick and loss are common, often anticipated and usually easily managed occurrences that have no connection with the rare disaster of the volcano. He did not wish to be named because he is currently under criminal investigation, along with a dozen other employees of EMP, Lapindo and subcontractors involved in the operation of the drilling project. "Kick and loss are like falling off a bike. It happens all the time. You just



The likelihood that an earthquake will trigger events such as mud volcanoes depends on its magnitude and distance. The black line shows the cut-off above which no connection is expected. In the case of East Java, some of the recent earthquakes that have not caused mud volcanoes (yellow dots) are below the cut-off, and some have been closer to Sidoarjo and stronger than the Yogyakarta earthquake of May 2006 (red dot).

get back on. What you want to avoid is getting hit by a truck. That's what happened the next morning," he says.

But Davies and Bachtiar think the 'truck' came up through the borehole. "They drilled something overpressured, something that was driving the system," Davies says. Bachtiar, one of the first people to suggest that the drilling might have been connected with the mud volcano, was a witness in a police investigation

into the possible link and thus studied Lapindo's drilling report closely. He suggests several ways in which what went on at the drilling site could have led to the eruption.

It is possible, he suggests, that the workers at the well withdrew their drill too quickly, losing control of the pressure in the hole. Istadi admits that they might have pulled out too fast, and that the effect could have been to suck in fluids from pressurized pockets in the rock, leading to the kick. But he insists that the kick was killed.

Soft spots

To 'kill' a kick, drillers circulate drilling fluid heavy enough to fight back the incoming liquid and gas. But this is a delicate process. The heavy fluid itself can open cracks in the surrounding rock, which the upcoming fluids can then enter. Lower strata are less vulnerable to cracking because the weight of overlying rock holds them together. In the shallower regions, steel casing is cemented to shore things up. The most vulnerable point is just below this casing — about a kilometre in the case of Banjar Panji-1. If the pressure caused by the heavy drilling fluid exceeds the pressure holding the rock together here, "it will be a disaster," says Bachtiar.

Rudi Rubiandini, a petroleum engineer at the Bandung Institute of Technology in Indonesia who led an investigation into the mud flow last June, believes that evidence in the drilling report shows that this is what happened. "It was very clear that the [rock] formation cracked," claims Rubiandini. The timing and the geographical location also suggest a connection between Lapindo's kick and the eruption.

A unique plan to stem the flow

Three Earth scientists at the Bandung Institute of Technology have hatched a plan to drop 1,000 chains, each with four concrete balls and weighing some 300 kilograms, into the neck of the mud volcano.

The aim is not to plug the hole, which could divert the pressure elsewhere, but to "tire" it out by making it work its way around the concrete balls. By restricting the size of the conduit, the team plans to cut the rate of flow (see *Nature* 445, 470; 2007). In the end, the same volume of mud will probably pump out, but at a lower rate. The innovators calculate that the rate of flow should fall by 75%,

giving them more time to reinforce embankments, channel the mud to the river and take other measures to spare the villages. "The balls will buy some breathing time," says Andang Bachtiar, an oil-exploration consultant at Jakarta-based Exploration Think Tank Indonesia.



Scientists have generally given it mixed, and many-caveated, reviews. "It does seem to me that it is basically reasonable to assume that there is some pressure distribution that is causing flow, and that partially blocking the conduit will lower the flow rate," says Steve Tait, a volcanologist at The Institute of Earth Physics in Paris, France.

But Tait adds: "One other outside possibility could be that if an outlet is blocked, the pressure distribution in the system creates a new conduit and point of emission."

Other scientists have noted, half in jest and half in fear, that putting concrete balls in the neck of a mud volcano could turn it into a cannon, shooting the balls straight back out.

Tait also points out that

predictions about how the experiment will work depend on knowledge of where the mud is coming from and what path it is taking to get to the surface — things about which little is known.

Bagus Endar Nurhandoko, one of the project's creators, admits that the project could run into problems straight away if they misjudge the shape of the crater and shoot the balls to the wrong place. "The worst thing would be if the balls don't go down at all," he says. They are using sonar to try to figure this out before inserting the balls.

And as Satria Bijaksana — another of the Bandung trio — admits, many things that work on paper do not in reality. "This is real life. You can't take into account all the parameters."

D.C.

According to Istadi, the kick was resolved within four hours. Moreover, he claims that, at 2 p.m. on 29 May, long after Lusi had started, Lapindo did a test that showed that the borehole was not fractured, at least not at 1,091 metres, the weakest point. Below that, he says, the borehole would have been strong enough to withstand the force of the kick. Bachtiar says, however, that such tests are open to interpretation; other experts contacted by *Nature* also wondered how such a test could have worked.

Mark Tingay, a geologist at the University of Adelaide in Australia, says the Sidoarjo volcano has a striking similarity to drilling-induced eruptions offshore from Brunei in 1974 and 1979 (M. R. P. Tingay *et al.* *J. Geol. Soc.* **162**, 39–49; 2005). There, deeply buried fluids under high pressure rose to a shallower rock formation that they then fractured, thus eventually reaching the surface. The event also showed the pattern of loss, kick and then eruptions seen in Lusi, some of which were kilometres from the drilling site. In the Brunei case, Shell, the company responsible for the drilling, has documented the expulsion and its efforts to alleviate the situation. The flow took more than 20 years and more than 20 relief wells to quell, says Tingay. “The similarities all suggest a man-made cause for Lusi,” he says.

Money matters

Indonesia's president, Susilo Bambang Yudhoyono, has already forced Lapindo to pay 3.8 trillion rupiah (US\$420 million) to help deal with the disaster, 2.5 trillion to provide relief and buy the despoiled land from its owners and 1.3 trillion to try to stem the flow. For Lapindo, the downside could get even worse if more villages flood or if it loses a lawsuit brought by Medco, Indonesia's largest oil producer and one of Lapindo's two partners in the joint venture responsible for the drilling.

Medco alleges that Lapindo should have put casing down to 2,591 metres to stabilize the hole, which might have made it possible to prevent or control any damage done by the kick. Some experts argue that the borehole should be cased for at least two-thirds of its depth and that Lapindo cut corners to save time and money. Lapindo, in a legal document sent to Medco on 2 February, countered the allegations by saying that the operating agreement requires Medco to cover its part of the liability. The letter ends: “Medco's ongoing failure to pay is jeopardizing the success of the Sidoarjo relief effort.” The Australian company Santos, which holds 18% of the joint venture and has several other drilling operations in Indonesia, had paid US\$15 million to the relief effort as of 8 December.

"If the Yogyakarta earthquake caused that mud volcano, it would have been way out of range."
— Michael Manga



Wanderer above the sea of mud.

People in Indonesia have already made up their minds. If “mud volcano” is mentioned, they shoot back, “Ah, lumpur Lapindo” — the Lapindo mud. And the locals do not think “Lapindo relief” is coming fast enough, although the company says that it has already spent extensively on medical assistance, food, emergency and temporary housing and cash allowances for the people of the villages.

Despite the damage, the area has become a spectacle as well as a wasteland. Locals have become self-designated toll masters trying to recoup from visitors what the Big Hole took from them. Making left turns, right turns and U-turns on the roads costs 1,000 rupiah (about 10 cents), as do roadside parking spots that had previously been mere dusty lots. Some bare-chested musicians on the side of one highway, who call themselves ‘Victims of the Hot Mud’, play guitar and dance to drums.

A motorcycle guide at the site is happy to take me along one of the earthworks. At one point, he kneels down and tosses a rock some 20 metres. It plops. “That is where my house is,” he says. Looking in the opposite direction from the same embankment, he points to a partially submerged factory where he used to make plastic and rattan. Lapindo, as part of a government order, has agreed to buy the land and property affected — but it is not clear when any payment will be made. In the meantime, my guide makes a living by guiding tourists to see the destruction that waylaid his home and job and selling them CD-ROMS of photography — although mine didn't work. As a motorcycle guide he makes 30,000 rupiah a day, compared with 700,000 rupiah a month in the old days. “Not bad,” I say. “I have no house,” he reminds me.

Meanwhile, there are some attempts to deal with the problem at the source. Rubiandini says that last year he and his team planned a series of relief wells to intercept the conduit bringing the mud up and staunch the flow by sending heavy mud back down. Two wells were started; both failed before reaching a third of the required 2,500–2,800 metres. “The drilling teams were constantly off and on because the money wasn't there. People are not serious about killing this,” he says, apparently referring to Lapindo. Istadi retorts, “We gave him \$80 million. We gave him everything. Even if he had had more time it would not have worked. Anyway, the whole project assumes that the borehole is to blame. And we don't know that.”

Energy zapper

The government has also undertaken a project to push the mud out into a nearby river, so far with little success. The latest hopes rest on an untested scheme to drop concrete balls down the Big Hole to soak up the energy of the upcoming flow (see ‘A unique plan to stem the flow’). But for now, it seems that the mud will continue to flow, and the ground will continue to sink under its weight. The hopes for the 200,000 houses under threat depend on the effective use of levees. Basuki says they have already brought in 2.5 million cubic metres of dirt. “We moved a mountain,” he says. But more science is needed too, since things might not merely continue, but get worse. For example, says Davies, the central vent could collapse as more and more mud gets removed from the subsurface and spreads its burden over the land. Other mud volcanoes have sunk 500 metres at the centre and forced land to sag for five kilometres around. The Big Hole could yet get bigger still.

David Cyranoski is *Nature's* Asia-Pacific correspondent.

D. CYRANOSKI



The flower of seduction

Hundreds of orchid species lure their pollinators with the promise of sex, only to send them away unfulfilled. **Heidi Ledford** looks at how dishonesty gives them the evolutionary edge.

He's a loner and, truth be told, he's never been with a lady before. She's an experienced siren with an irresistible scent. Lightning quick, he zigzags close, strokes her hair and mounts her. But it is only as he engages in a few confused attempts at copulation that the truth dawns: his chosen is not a lady after all. She's a flower.

So goes the tale of the male solitary bee (*Andrena nigroaenea*) and his seducer, the green and dainty early spider orchid (*Ophrys sphegodes*). These orchids bloom in the spring, when female bees are still in their nests and the males are young, inexperienced, and likely to encounter an orchid before they find a real mate. During his attempts to mate with the elaborate flower, the bee may pick up a large packet of sticky pollen on his head, ready to deliver to another plant. That's a boon to the orchid; the bee, on the other hand, gets nothing but hot and bothered.

Some 30,000 species strong, orchids are one of the largest groups of flowering plants; collectors covet their elaborate and sometimes rude-looking flowers. But the plants are also unusual in their dishonesty. Although most insect-pollinated plants pay their pollinators in energy-rich nectar, about a third of orchids offer no rewards. Roughly 10,000 species deceive their pollinators by mimicking plants that do provide nectar. And, unique among plants, another 400, including the early spider orchid, mimic females and promise their pollinators sex.

Researchers have recently dissected the

scents that some deceptive orchids use to lure insects. Because each species manufactures a unique cocktail, and typically attracts only one species of pollinator, the odours provide a way for biologists to identify species in this class of orchid and to study how they have diversified. These studies may eventually provide evidence for sympatric speciation, the concept that a new species can evolve without being geographically isolated. Indeed, orchids may be one of the most promising ways in which to find evidence for this controversial idea.

At first glance, dishonesty seems to be a bad policy for orchids. Non-rewarding flowers have low pollination rates, and produce only half as much fruit as rewarding flowers¹. Add nectar to a non-rewarding plant and pollination rates shoot up. If reproductive success was measured by seed production alone, then deceptive orchids would have lost the evolutionary race long ago.

But deception may be worthwhile because it prevents inbreeding and so boosts the quality, rather than quantity, of seeds. Deceived pollinators are more likely to go off in a huff after realizing they've been tricked: they'll take their

package of pollen and fly to a more distant patch of flowers, reducing the chance of landing on a close relative of the original plant². "To be deceptive means that the orchids have less sex, but the sex is better because it's not with a close relative," says Salvatore Cozzolino, an evolutionary biologist at the University of Naples in Italy.

A few years ago, Florian Schiestl, a biologist at the Swiss Federal Institute of Technol-

ogy in Zurich, set out to determine just what makes orchids masquerading as females so attractive to insects. Although the flowers do look and feel a bit like a female — the hairs on *O. sphegodes* flowers, for example, resemble those on a female bee — Schiestl knew the key to long-distance attraction was to smell like a female, because bees respond more to distant odours than to visual cues. But plants produce hundreds of volatile compounds to repel predators and microbes, so the precise perfume was difficult to determine. "You don't know which are important for the pollinators and which are not," Schiestl says.

Sexy smell

Schiestl and his colleague Wittko Francke of the University of Hamburg in Germany reasoned that the way to intoxicate a bee must be through its antennae. Bee antennae are carpeted with hundreds of sensors that carry receptors to detect odours and dispatch electrical signals to the brain. To figure out which floral compounds were attracting male pollinators, the researchers extracted compounds from the labellum of the flower, where insects land, and then separated the chemicals using gas chromatography. They plucked a bee antenna, placed it between two electrodes and exposed it to each component of the mix. When a compound bound a receptor on the antenna, a tiny current passed through the apparatus.

Using this method, Schiestl and Francke were able to show that the complex perfume that arouses male *A. nigroaenea* bees is made up of 14 different compounds that are also common components of the waxy cuticle

"To be deceptive means that the orchids have less sex, but the sex is better because it's not with a close relative."

— Salvatore Cozzolino

that protects the surface of many plants³. They showed that the same combination of compounds is present in the volatile sex pheromone that a female bee uses to attract a mate, and that a blend of these chemicals could make bees mate with dummy flowers. The finding also revealed how sexual deception could have evolved in this species by gradual modification of systems the plant was already using to make its own compounds. Each tweak in the ratio of compounds that increased pollinator visitation would have given the orchid a reproductive advantage.

The discovery also suggested one way in which orchids could evolve new species. Most insects are very picky about the scent they respond to and often pollinate only one orchid species. So if a plant changed its aromatic bouquet enough it would attract a different pollinator and rebuff the old one. Over time, that would — theoretically at least — result in genetic isolation and the creation of a new orchid species. “We think that speciation can occur fairly quickly in that system,” says Schiestl. “The plants need only to change their odour bouquet to attract a new pollinator.”

The team found evidence to back this idea in the orchid blooms of Australia. They repeated the experiment on the orchid *Chiloglottis trapeziformis*, which tricks the male thynnine wasp (*Neozeleboria cryptoides*). Analysis of *C. trapeziformis* scent revealed a surprise — rather than adapting existing mechanisms, the orchid was producing an entirely different chemical compound they named chiloglottone, which is also a pheromone made by female wasps⁴. They also found that another *Ophrys* species, *O. speculum*, concocts a different wasp pheromone by developing several novel compounds⁵. In this case, the orchid mimic worked so well that, when offered a choice between a female or an orchid, male wasps courted the orchid.



R. PEAKALL

Some species of *Chiloglottis* orchids look identical but use different chemicals to attract pollinators.

Rod Peakall, an evolutionary biologist at Australian National University in Canberra and his colleagues have now surveyed 10 of the 30 known species of *Chiloglottis* orchids across Australia. In some regions, the team found orchids growing together that were identical in appearance, location and flowering time yet were, based on genetic markers, not interbreeding and that belonged to different species. Invariably, flowers of the two species produced slightly different pheromones: one, for example, produced chiloglottone alone whereas the other mixed chiloglottone and a slightly modified version. The two pheromones attract different species of wasp, and targeting different pollinators seemed to have defined species boundaries by preventing the flow of genetic information between the two sets of orchids.

Growing apart

If Peakall can prove that is the case, he may have evidence for sympatric speciation. This has been notoriously difficult to confirm because, in many systems, it is tough to prove that two similar species growing together were not spatially separated in the past and then reunited. The few documented examples have been in geographically remote areas, such as isolated islands or crater lakes, where the influx of species from other regions is unlikely. “Having solid evidence of sympatric speciation in a continental setting would be a big deal,” says Daniel Bolnick, an evolutionary biologist at the University of Texas, Austin, “but it would have to be very strong evidence.”

Peakall plans to seek such evidence by using small regions of highly variable genome sequences to track the evolutionary history of his orchids. The aim is to show that the two species growing together are more closely related to one another than either is to orchids

located farther away. To do this, he must piece together a precise orchid family tree and rule out the possibility that two similar species actually evolved because they were once geographically separated and later came to grow side by side again. Cozzolino warns that this will be tough because the DNA sequences of these orchids do not vary by much and it may be difficult to distinguish between very closely related species. Nevertheless, Cozzolino agrees that sympatric speciation may be under way, and is looking for evidence of it among European sexually deceptive orchids as well.

And what of the insects in this web of deception? Both orchid and insect enter into the encounter in the hope of better sex — but only the flower comes away with it. This apparent inequality is such that Charles Darwin refused to believe that pollinators would waste their efforts on orchids that offer no reward. He expected the insects to figure out the ruse and save energy by avoiding the flowers and forcing the orchid into extinction.

But evolutionary biologists now know that young insects gain more by enthusiastic but indiscriminate mating than they would by being more choosy. Females are often in short supply and there is intense competition among males to find a mate — so it is better for males to try and fail than not to try at all. “The males aren’t too picky,” Peakall says, “Their strategy is ‘Hey, I will go for anything that looks like a female because I can’t afford not to.’”

Heidi Ledford writes for *Nature* from Berkeley, California.

1. Tremblay, R. L., Ackerman, J. D., Zimmerman, J. K. & Calvo, R. N. *Biol. J. Linnean Soc.* **84**, 1–54 (2005).
2. Johnson, S. D., Craig, P. I. & Ågren, J. *Proc. R. Soc. Lond. B* **271**, 803–809 (2004).
3. Schiestl, F. P. *et al. Nature* **399**, 421–422 (1999).
4. Schiestl, F. P. *et al. Science* **302**, 437–438 (2003).
5. Ayasse, M., Schiestl, F. P., Paulus, H. F., Ibarra, F. & Francke, W. *Proc. R. Soc. Lond. B* **270**, 517–522 (2003).

F. P. SCHIESTL



Faking it: the early spider orchid lures male bees to its pollen by mimicking the female sex pheromone.

Concrete evidence that volcanoes can be stopped

SIR — In your News story “Volcano gets choke chains to slow mud” (*Nature* **445**, 470; 2007), you cite some physicists who had not heard of using concrete balls and linked chain to attempt to plug a mud geyser. The US Navy (in which I was then a commander), the Italian Navy and the US Marine Corps used exactly this approach successfully in Sicily, in April–May 1992, to slow and eventually redirect lava flow from Mount Etna that was threatening the small village of Zafferana.

We came up with this plan while talking to an Italian geologist. He spoke no English and I spoke no Italian, but I have a bachelor's degree in geology and that helped a lot. We communicated on the back of a napkin while seated at a small restaurant at the ski lodge that became our base of operations. The navy units involved tried several different ways to place large decommissioned anti-terrorist barriers into a vent approximately 8,000 feet (2,440 metres) up the side of the volcano.

The first plan was to drop individual barriers into the vent, but that failed because of insufficient quantity, and the heat simply ignited the concrete. We then built a very large slide and were going to stack barriers on the slide and slip it into the vent. The slide was constructed but was impossible to move to the vent as mountain winds forced it into the path of the aircraft. (I filmed it on video and it is rather dramatic.)

Our third plan, which eventually worked, involved linking several dozen barriers together with asbestos-wrapped anchor chain and placing them in position around the vent. Additional concrete ‘Dempster Dumpster’ pads were placed directly over the vent on a net formed of anchor chain. When the entire assembly was in place it was blown into the vent using plastic explosive. US Navy and US Marine Corps CH-53E helicopters were used to place the barriers and transport the Italian explosives ordinance team required to place the explosives. The lava tube carrying lava down the mountain to the vicinity of Zafferana collapsed when flow was interrupted. That removed the immediate danger.

The long-term solution was to drive a bulldozer up the mountain, dig a very large canal and blow out the side of the vent. We had to change the engine on the bulldozer when we finally got it in place, as moving it up the hill destroyed the original engine.

The operation was extremely hazardous and at times conducted in blinding snowstorms. Aircrews had to contend with the noxious mix of gases as they hovered directly over the vents for long periods of time. Several US Navy and US Marine Corps aircrews were awarded air medals for their performance. The entire operation was

designed, developed and coordinated by the Air Operation Department at Naval Air Station Sigonella, Sicily, under the command of Captain Michael Bruner. I had the privilege of running the Air Ops Department during the Etna operation and have 35 air crew hours logged in the CH-53Es operating on the mountain.

John Carpenter

Annandale, Virginia 22003, USA

See News Feature, page 812.

Newspaper scare headlines can be counter-productive

SIR — Your coverage of the report from the Intergovernmental Panel on Climate Change (IPCC) working group included some exemplary Editorial and News headlines: “Light at the end of the tunnel”, “What we don't know about climate change” and “From words to action” (*Nature* **445**, 567 & 578–583; 2007). These convey the message about knowledge, ignorance and action that would be expected from a leading journal writing for a scientific readership.

Communicating science to wider, public audiences, however — in this case on matters of important public policy — is an art that requires careful message management and tone setting. It seems that confident and salient science, as presented by the IPCC, may be received by the public in non-productive ways, depending on the intervening media.

With this in mind, I examined the coverage of the IPCC report in the ten main national UK newspapers for Saturday 3 February, the day after the report was released. Only one newspaper failed to run at least one story on the report (one newspaper ran seven stories), but what was most striking was the tone.

The four UK ‘quality’ newspapers all ran front-page headlines conveying a message of rising anxiety: “Final warning”, “Worse than we thought”, “New fears on climate raise heat on leaders” and “Only man can stop climate disaster”. And all nine newspapers introduced one or more of the adjectives “catastrophic”, “shocking”, “terrifying” or “devastating” in their various qualifications of climate change. Yet none of these words exist in the report, nor were they used in the scientists’ presentations in Paris. Added to the front-page vocabulary of “final”, “fears”, “worse” and “disaster”, they offer an insight into the likely response of the 20 million Britons who read these newspapers.

In contrast, an online search of some leading newspapers in the United States suggests a different media discourse. Thus, on the same day, one finds these headlines: “UN climate panel says warming is man-made”, “New tack on global warming”, “Warming report builds support for action”

and “The basics: ever firmer statements on global warming”. This suggests a more neutral representation in the United States of the IPCC's key message, and a tone that facilitates a less loaded or frenzied debate about options for action.

Campaigners, media and some scientists seem to be appealing to fear in order to generate a sense of urgency. If they want to engage the public in responding to climate change, this is unreliable at best and counter-productive at worst. As Susanne Moser and Lisa Dilling point out in *Creating a Climate for Change: Communicating Climate Change and Facilitating Social Change* (Cambridge Univ. Press, 2007), such appeals often lead to denial, paralysis, apathy or even perverse reactive behaviour.

The journey from producing confident assessments of scientific knowledge to a destination of induced social change is a tortuous one, fraught with dangers and many blind alleys. The challenging policy choices that lie ahead will not be well served by the type of loaded reporting of science seen in the UK media described above.

Mike Hulme

Tyndall Centre, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Niche drugs aren't a cheap alternative to blockbusters

SIR — Your Editorial “A changing drug supply” (*Nature* **445**, 460; 2007) captures some of the realities facing the pharmaceutical industry in the post-blockbuster era. However, tailored drugs with smaller targeted populations are not necessarily cheaper to develop. If this were true, the pharmaceutical industry would have migrated to this model long ago.

Drugs targeted at smaller populations still need to be discovered, validated in pre-clinical models, formulated, and shown to be efficacious and safe in patients — in short, the same requirements as for a potential blockbuster drug. Depending on the disease and a multitude of factors, some clinical trials could potentially be less expensive for a small population, but this is unlikely to be the norm. Going down this path does not guarantee a lack of competition either: as revenue pressures build, more companies will compete for these niche populations. Targeting small populations is not a panacea for the industry illness, just as there are none for human illness.

W. Ross Tracey

11 North Ledge Rock Road, Niantic, Connecticut 06357, USA

Contributions to Correspondence may be submitted to corres@nature.com. Published contributions are edited.

BOOKS & ARTS

Calamity gene

When biotechnology spins out of control.

Next

by Michael Crichton

HarperCollins: 2006. 448 pp.

\$27.95, £17.95

Michael A. Goldman

At last, a Michael Crichton novel you can put down, a vilification of science that does not make compelling reading. *Next* is a veritable catalogue of what could go wrong with biotechnology. This is what would happen if every patent attorney and judge had a pre-frontal lobotomy (done the humane way using RNA interference, of course), in case the bioethicists hadn't warned you already. Crichton tries to address every aspect of the biotechnology craze at once, giving the book too many simultaneous plotlines to follow. The result is no single story you really want to stick with to the end.

The book is certainly entertaining at times. There is a smart-beaked African grey parrot named Gerard, rather self-centred and with a British accent. In exile because he imitated his owner's husband in a liaison with another woman in the house, Gerard manages to help a chimpanzee with a raspy voice and bad grammar rescue an ordinary little kid from being abducted for his liver. Meanwhile, a cocaine-addicted good-for-nothing thirty-something finally snorts a concoction that turns his life around: an experimental virus containing a human 'maturity gene'. The transgenic orangutan who curses in Dutch could use a whiff of that stuff as well. Body parts and eggs are traded like commodities, and genetically engineered fish display illuminated advertising billboards on their sides. I admit that Crichton had me going with the *Tyrannosaurus rex* in *Jurassic Park*, but somehow that was a bit more believable, and had a lot more suspense.

One story line concerns BioGen's ownership of Frank Burnet's cell line, which produces a cancer-fighting compound. BioGen gets a legal judgement that says they own Burnet's cells, wherever those cells happen to be. When BioGen's own lines are contaminated, and Burnet himself has disappeared, the company employs bounty hunters to collect the 'property' from Burnet's daughter and grandson. Dave, the genetically enhanced chimpanzee, effects a rescue. The Burnet family's attorney challenges the property decision. After an exasperating hearing, a judge finally hands down a reasonable decision — this is the first time that *Next*



The big screen: the film *Gattaca* portrayed a future in which opportunities are dictated by genetic tests.

seems like anything other than an unabashed condemnation of biotechnology. Human cells, the judge says, are different from other forms of property, as one might still be attached to them, especially when they are in an unwilling body.

Crichton uses extant biotechnology company names, such as BioGen and BioRad, even if he does capitalize them differently and change the trademarks. Even California senator Dianne Feinstein plays a part. Craig Venter seems to have escaped, but there's a spiritual and influential scientist at the National Institutes of Health associated with the thrill-seeking gene. And a prominent divorce attorney to the stars discovers the value of genetic testing — for everything — in determining whether or not a parent is a good candidate for child custody.

As in his previous novel, *State of Fear* (HarperCollins, 2004), Crichton pleads his case clearly and succinctly in an author's note at the end of the book. He lists five points, which I take up in turn.

First, Crichton contends that we should "stop patenting genes" because vague claims render it impossible to do useful research and paralyse scientific progress. Although this idea has been bandied about, there are powerful contrary arguments. Patent protection actually encourages the publication of results, and

requires a full description of an invention so that it can be used by others. Patent enforcement typically does not occur until a competitor is ready to market a commercial product, at which time a user licence can be negotiated; basic research is not excluded. Crichton's claim that a patent on 'noses' would exact a licence fee from perfume manufacturers is too broad an interpretation; a patent on a gene clearly does not even extend to a drug that inhibits the activity of the product of that gene. We do, however, have a problem if patents are granted too broadly and on too little information, relying strictly on enforcement actions to ascertain validity.

His next claim is that we should "establish clear guidelines for the use of human tissues". In *Next*, and in a commentary in *The Wall Street Journal* on 15 December 2006, Crichton describes a case in which a court upheld the right of Washington University to keep possession of prostate-cancer tissues donated by patients of William Catalona when he moved to Northwestern University. The ruling was astounding in that it ignored the wishes of 6,000 patients who had expressly consented to the move, thereby asserting that patients had no control over the tissues once they had left the body. This apparently contradicted the patient consent forms, which said that patients could withdraw from the study at any time, and

COLUMBIA/KOBAL COLLECTION

that tissues would be used only for the specified purpose. Crichton makes an important point about the Catalonia case, but the portrayal of this issue in *Next* is extreme.

Crichton thinks we should “rescind the Bayh–Dole Act”, which allows US universities to claim ownership of patents derived in whole or in part from federally funded research done in the academic setting. He believes that this 1980 act has turned universities into money-hungry corporate monsters no longer motivated to seek truth, but rather to prove the efficacy of the pharmaceuticals their research labs produce. Thus, he contends, university faculty can no longer be relied on for an unbiased opinion on anything, let alone a reliable piece of data. The jury is still out on the Bayh–Dole act, but academic institutions and scientists would probably be looking beyond tuition payments, gifts and federal grants for funding with or without Bayh–Dole.

Most readers of *Nature* would surely agree

with Crichton's other claims, that we should “pass laws to ensure that data about gene [therapy] testing is made public” and “avoid bans on research”.

The book could have a role as a conversation starter for a course in bioethics. Most students would find it more entertaining than the typical textbook, and it covers a similar range of issues. There is no doubt that our scientific capabilities and our imagination have already gone well beyond the reaches of law and ethics, even though *Next* exaggerates the extent. The book is a reminder that we need to educate professionals in medicine and law about biotechnology so that the field's gargantuan potential, which Crichton's book never denies, can be realized safely and equitably. It never hurts to have this highlighted on *The New York Times* bestseller list. ■

Michael A. Goldman is in the Department of Biology, San Francisco State University, San Francisco, California 94132-1722, USA.

In Western culture there is an abiding distinction between understanding nature and doing things with it. This was formulated by Aristotle as *episteme* versus *techne* (the Latin equivalents are *scientia* and *ars*). Natural philosophy, the discipline responsible for seeking a causal understanding of natural qualities and processes, was classed as *scientia*, whereas mathematics, which deals with quantities that need not apply to real things, was regarded as a practical craft charged with measurement and computation. As Dear stresses, Isaac Newton's universal law of gravitation — the great scientific achievement of the seventeenth century — was viewed as a mathematical accomplishment. Although highly sophisticated, it was in the same class as utilitarian calculations of the relative movements of the stars and planets involved in almanacs and horoscopes. Moreover, Newton eschewed any account of what gravity was or how it acted at a distance in a vacuum. He did not provide a natural-philosophical explanation of the kind attempted by René Descartes, for example, who sought to understand celestial motion in terms of bodies in a fluid medium using the analogy of straws in the eddy of a river. Such an account was deemed intelligible, even if not demonstratively true. Newton's reluctance meant that his theory lacked intelligibility, even though it possessed striking instrumentality, as judged by the predictive power of the inverse-square law.

In discussing various scientific domains from the seventeenth century to the present, including celestial mechanics, taxonomy, atomism, natural selection, electromagnetism and quantum physics, Dear meditates on this

The two faces of science

The Intelligibility of Nature: How Science Makes Sense of the World

by Peter Dear

University of Chicago Press: 2006. 242 pp. \$27.50

Richard Yeo

In 1690 the philosopher John Locke imagined a man with “microscopical eyes” many times more acute than the best microscope. Such a man, he conjectured, might grasp the deep “texture and the motion of the minute parts of corporeal things”, but “would be in a quite different world from other people... I doubt, whether he, and the rest of men, could discourse concerning the objects of sight.” We can surmise that the scientific explanations such a person offered might not be intelligible to others.

Intelligibility may be difficult to define but it plays a crucial role in the claim of science to offer credible accounts of nature, argues Peter Dear in his elegant book *The Intelligibility of Nature*, which is richly informed by scholarship in the history of science. Locke's point looks prescient when we consider the development of quantum mechanics from the 1920s. Its successful predictions of experimental phenomena were aligned with the disorienting prospect of an acausal, probabilistic world. Dear argues that by this time, ‘instrumentality’ — the power to produce and predict effects — had surpassed intelligibility

as the main basis of scientific authority. The leitmotif of his book is that science has availed itself of two self-supporting, albeit circular, rationales: its account of the structure and processes of nature is backed by the success of instrumental techniques (such as the use of electron microscopes and DNA profiling); and explanations of why certain techniques work are grounded in intelligible, even if speculative, accounts of the natural world.



Isaac Newton (left) and Antoine-Laurent Lavoisier were central to the creation of modern science.

tension within science. He contends that Newton's success and reputation allowed "the conflation of natural philosophy with instrumentality". Antoine-Laurent Lavoisier reinforced this by insisting that precise measurement of chemical reactions could be a model of scientific method without any commitment to a view about the underlying elementary structure of matter. In contrast, John Dalton's atomic theory was an attempt at traditional natural philosophy. In so far as Lavoisier was triumphant, the criterion of instrumentality began to rival that of intelligibility, the hallmark of natural philosophy. By the nineteenth

century, modern science "was born a hybrid of two formerly distinct endeavours".

In 1833, William Whewell, surprisingly absent from the book, coined the term 'scientist' in opposition to 'artist'. Both the date and the terms of this contrast resonate with Dear's theme. If the dynamic between the two rationales for science was already in play by Newton's day, what was added in the nineteenth century? The dust-jacket suggests one answer, declaring that the investigation of nature "would be carried out by a new kind of person, the scientist". Although the features of this new persona are not made explicit, there are indications that it

was, and remains, a problematic one. For a cosmologist, being a scientist might be a natural-philosophic quest for laws of nature that cannot become intellectual property; for a microbiologist, it is likely to involve a search for techniques that will be patented. Scientists who wish to reflect on their vocation will gain valuable insights from this beautifully contrived book, and all readers will be prompted to think more carefully about the nature and ethos of science. ■

Richard Yeo is at the Centre for Public Culture and Ideas, Griffith University, Nathan, Queensland 4111, Australia.

Ordering according to size

Why Size Matters: From Bacteria to Blue Whales

by John Tyler Bonner

Princeton University Press: 2006. 176 pp.
\$16.95, £9.95

Victor Smetacek

I used to think that the story of the mother who starts a letter to her grown-up son with: "Dear son, I am writing this letter slowly because I know you cannot read fast," was just funny. But after reading the tiny book *Why Size Matters* by John Tyler Bonner twice, I now realize that the story had a deeper meaning than had first met my eye. Bonner says in the preface that the big picture he has painted on a small canvas took a lifetime to mature and that writing this book was a slow process. Indeed, I found myself reading it slowly as well, making sure that I did not miss anything. The territory covered is so vast, as indicated in the subtitle, that I often stopped to get my bearings and ponder the view.

The message of this book is that size rules biology to a greater extent than most of us are aware, because our untrained mind's eye expects its sense of proportion to be universal, whereas in reality, it is relative. When things get bigger, their length increases linearly but their surface area increases by the square and their volume by the cube. So when the proportions, properties and performances of organisms, such as body shape, life span or speed of movement, are plotted against size or weight in log-log graphs, they fall on straight lines with interesting exponents.

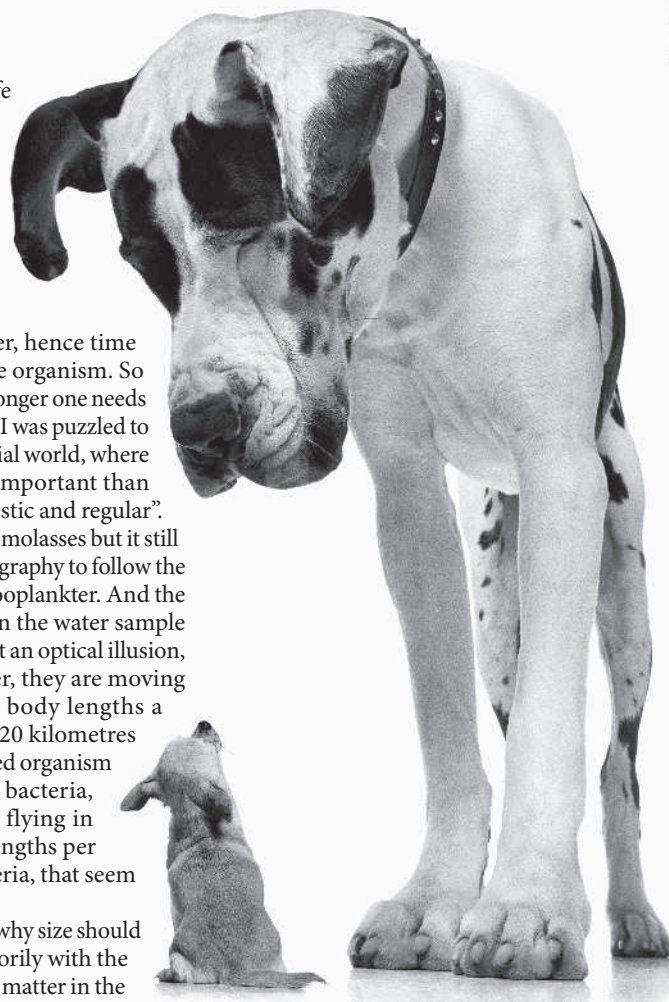
The relationship between weight and strength dates back to Galileo, who calculated the effect of increases in animal size on body proportion. I find his sketch comparing the bones of a 'normal' and an oversized animal greatly exaggerated. I cannot train my mind's eye to conjure up images of small and large animals commensurate with Galileo's bones. To me, a tiger viewed from a distance looks as graceful as a house cat. On the other hand, the branch-breaking, heavy-duty elephant and the

leaf-nibbling, slender giraffe look so different because they do different things, which need varying levels of strength. Yet both animals need to spend most of the day feeding. But these are minor details on log-log graphs.

In general, life is shorter, hence time flies faster, the smaller the organism. So the bigger the picture, the longer one needs to get it. Perhaps that's why I was puzzled to read that life in the microbial world, where viscosity becomes more important than gravity, is "very slow, majestic and regular". The medium might be like molasses but it still takes high-speed cinematography to follow the movements of a feeding zooplankton. And the bacteria zipping around in the water sample under a microscope are not an optical illusion, as Bonner suggests; rather, they are moving at more than a hundred body lengths a second — equivalent to 720 kilometres per hour for a 2-metre-sized organism like humans. It is not the bacteria, but the planes overhead, flying in the same range of body lengths per second as the fastest bacteria, that seem to move majestically.

Bonner argues at length why size should matter, but deals perfunctorily with the question: 'What made size matter in the first place?' Selection by predators is the most obvious reason, with increasing size a type of escape. A 'food-for-thought' insight I had was the statement 'there is always room at the top', implying that organisms can always get bigger. But the bigger they are, they harder they fall, so it is the small ones that are less vulnerable in the long run. There is another intriguing insight on size and speed waiting for the reader in the penultimate chapter, but I will not divulge it here.

Bonner's style is to open windows and then step aside so that his readers can see for themselves. If you are up to date with the recent



Size differences among animals can result in great differences in the way they move and live.

literature on size rules then you have already looked through many of these windows, but if you have not done so consciously, or if you are studying the topic, this pithy little book is worth the time spent reading it, whether slowly or not. ■

Victor Smetacek is professor of bio-oceanography at the University of Bremen, and at the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven, Germany.

For the love of Greece

King Ludwig I of Bavaria was a lifelong Hellenophile whose son Otto became king of Greece in 1832 after the country was freed from Turkish rule. Ludwig commissioned Carl Rottmann to produce 23 paintings celebrating contemporary and ancient Greece for his palace gardens in Munich.

The exhibition 'Ten Tons Hellas', to run at the New Pinakothek in Munich until 29 April, brings together 14 of these works — each weighing 400 kg — for the first time since the Second World War. But scientific examination of the paintings has revealed that Rottmann did not blindly follow the dictates of his powerful king.

The commission had specified that Rottmann should work with encaustic, a newly revived technique involving the mixing of pigments with melted wax or resins. Spectacular examples of

encaustic murals had recently been unearthed at Pompeii, their dazzling colours barely touched by centuries of winter rain. The encaustic technique was laborious and risky, involving heating at different stages of painting to

melt the resins. And because its revival was guided by the rather ambiguous writings of ancient scholars such as Pliny, it was still rather experimental.

Researchers at the Doerner Institute, housed in the New Pinakothek, examined some of the paintings. They found that by the time he was on the third painting,

Rottmann was experimenting with different methods to refine the encaustic technique. He soon abandoned all pretence and worked with a more conventional style of oil painting.

Ludwig had also dictated that the works be painted on moveable slabs. The researchers' X-ray analysis revealed welded iron support frames (shown here) — a technology developed for the railway industry — incorporating a network of wires to hold the layers of mortar and plaster. After the paintings were completed, they were set into the walls of a gallery in the New Pinakothek when it opened in 1853. Being so heavy, they remained in the gallery while other works were removed for safekeeping during the war. Stored in a basement, they were badly damaged by bombs in 1944.

In the new display, the paintings are once again mounted into the plaster walls to give the intended character of murals.

A.A.



A. ABBOTT

The pleasure principle

The Science of Orgasm

by Barry R. Komisaruk, Carlos Beyer-Flores & Beverly Whipple

Johns Hopkins University Press: 2006.
280 pp. \$25, £16.50

Tim D. Spector

The three academic authors of *The Science of Orgasm* — a neurophysiologist, a hormone physiologist and a sexologist — promise in the preface to take a different perspective from recent books exploring sexual behaviour and orgasm. Instead they focus on the physiology and pathology of ejaculation and orgasm in the context of bodily changes, health implications, dysfunction, ageing, pleasure and the nervous system. The publisher's blurb states that the book, which includes a glossary, sporadic definitions in the text and a meaty 45 pages of references, is accessible to lay readers.

The result is a bit of a pot-pourri. Although good in parts, the book lacks continuity and the authors do justice to only a few of their stated areas. They pose major problems for themselves by trying to cover too much and targeting both academics and lay readers — they certainly fail to deliver for the latter. Another problem is their loose definition of 'orgasm'. About a third of the book is related to sexual drive and desire, which includes much work on rodents and primates. These animals clearly ejaculate, but there is little supporting evidence (with the exception of a few primates)

that they orgasm as humans do. Confusingly, the authors usually consider male and female orgasm together and seem to assume, based on a single study, that male and female orgasms are indistinguishable.

The title *The Science of Orgasm* sounds as if it should include all the key facets. I am clearly biased — but do genetic influences warrant only three lines in 300-odd pages? They also omit any real mention of human variation. Why might only one in three women reach orgasm regularly, and why do so few men have anorgasmia? Other female topics surprisingly given little space are the contraceptive pill, hormone replacement therapy and placebo effects on orgasms, as well as potential new treatments for sexual dysfunction, such as PDE5 inhibitors, oxytocin sprays and testosterone patches. Even more odd is the lack of discussion of the science around the existence of the G-spot and female ejaculation — particularly as one of the authors, Beverly Whipple, was one of the G-spot pioneers in the 1980s.

The book gets bogged down in lengthy descriptions of hormone action, neuropharmacology and chronic diseases that read like a dull, albeit important, student textbook. There are some useful sections on the surprising and apparently real benefits of herbal remedies, such as ginseng, ginkgo and mixtures such as ArginMax. But the book only comes alive after about 200 pages with anecdotal stories of subjects who, after surgery, electrostimulation

and epileptic seizures, report having orgasms in their feet, while asleep or while visiting the physiology lab.

In my view, the best part of the book is the neurochemistry of the orgasm. Studies of paraplegic women clearly show the importance in female orgasm of multiple complex neural pathways such as the vagus nerve. Functional brain imaging is an exciting area for study and (despite poor-quality pictures) the authors present the latest findings of multiple areas of brain activity during orgasm — which make any simplistic dopamine (stimulatory)–serotonin (inhibitory) mode of action unlikely. They postulate a central role for areas such as the cingulate cortex, which is also where pain is perceived — linking pain and orgasm as related sensory processes. Orgasms apparently alter pain perception and increase pain thresholds, and this link may explain bizarre reports of women having orgasms during childbirth. However, just when I was ready for the truth — a clear definition of orgasm and where it arises in the brain — I was told it was not a reflex, only a perception of neural activity and, even worse, probably a form of diffuse consciousness in an as yet undiscovered fifth dimension. After such a careful, slow build-up of teasing and tantalizing data, I was definitely left frustrated — and wanting more.

Tim D. Spector is professor of genetic epidemiology, King's College London, London SE1 7EH, UK.



Bringing cartoons to life

To understand cells as dynamic systems, mathematical tools are needed to fill the gap between molecular interactions and physiological consequences.

John J. Tyson

Open any issue of *Nature* and you will find a diagram illustrating the molecular interactions purported to underlie some behaviour of a living cell. The accompanying text explains how the link between molecules and behaviour is thought to be made. For the simplest connections, such stories may be convincing, but as the mechanisms become more complex, intuitive explanations become more error prone and harder to believe.

A better way to build bridges from molecular biology to cell physiology is to recognize that a network of interacting genes and proteins is a dynamic system evolving in space and time according to fundamental laws of reaction, diffusion and transport. These laws govern how a regulatory network, confronted by any set of stimuli, determines the appropriate response of a cell. This information-processing system can be described in precise mathematical terms, and the resulting equations can be analysed and simulated to provide reliable, testable accounts of the molecular control of cell behaviour. To make these ideas clear, I will use a simplified example.

Sometimes the best response a cell can make is suicide. Programmed cell death involves activation of proteases, called caspases, that disassemble the doomed cell in a tidy fashion, avoiding the inflammatory consequences of cell breakdown. Cells stockpile caspases in precursor forms and activate the precursors under carefully regulated conditions. The network that activates caspase is extremely complex, with many levels of control to keep caspase activity switched off in healthy cells and to turn it on in cells that need to die.

Various sets of conditions can activate cell death. When the control system senses such conditions, it must trigger death irreversibly — the suicide command should not be withdrawn, even if the initiating signals are cancelled. In addition, the control system must not respond to inevitable pro-death fluctuations that are beneath the threshold for commitment to programmed cell death. These essential properties of cell death suggest that its molecular regulatory network is bistable (either off or on) at zero signal strength and monostable (on) for signals above the threshold.

Bistable behaviour of a chemical reaction system typically results from positive feedback, of which there are several

possibilities in the programmed-cell-death network. For instance, small amounts of caspase in the cell are neutralized by binding to an inhibitory protein XIAP — whose function is to prevent accidental firing of the suicide pathway by inadvertent activation of a few caspase molecules. But if enough caspase is activated

to saturate the XIAP pool, then the excess caspase triggers release of a protein whose job is to eliminate XIAP and free up even more caspase. These sorts of interactions might generate the kind of dynamic responses so characteristic of programmed cell death. But can we be sure our intuition is correct? Under what conditions is

the off state stable to small signals but switched irreversibly to the on state by large signals? How might the regulatory system fail? What are the most effective ways to intervene pharmaceutically to repair the cell-death pathway?

In the face of such quantitative questions about a complex molecular regulatory system, our intuition needs some guidance. The network of biochemical reactions among caspases, inhibitors, activators, inhibitors-of-inhibitors and so on, can be cast into a set of kinetic equations describing how the rates of the reactions cause the concentrations of the network components to increase or decrease in time. The 'state' of the control system can be represented as a point in a multidimensional coordinate system, and the kinetic equations define little vectors in this space, telling how every component of the control system will change over the next small increment of time. By following the arrows, a computer can simulate the temporal evolution of the control system under any specified experimental conditions. Computer simulations can then be compared with experimental observations to judge the adequacy (or inadequacy) of hypotheses based on the molecular interactions in the network.

More importantly, for nonlinear differential equations, a well-developed mathematical theory describes qualitative properties of equation solutions, and these properties accord well with our intuitive

notions; for example, 'bifurcation points' correspond to thresholds. Bifurcation analysis is a powerful tool for deducing qualitative dynamical features of complex reaction networks. In this fashion, dynamical systems theory forges a rigorous chain of deductions from molecular interactions to kinetic equations

to vector fields to physiological consequences.

This 'dynamical perspective' has proven its merits in many areas of molecular cell biology. Calcium signalling shows a variety of fascinating behaviours, none of which can be understood in quantitative detail without mathematical models. The molecular basis of circadian rhythms

is another area in which mathematical modelling is essential

to understanding such physiologically significant features as spontaneous oscillations, entrainment, temperature compensation and jet lag. In the study of cell-cycle regulation, bifurcation theory has led to novel predictions of excitation thresholds, time lags, checkpoint delays and oscillatory damping, all of which have been confirmed experimentally.

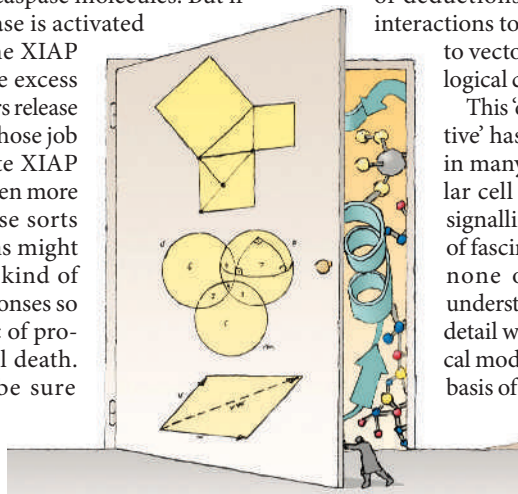
The dynamical perspective promises to revolutionize how we think about the molecular basis of cell physiology. For instance, the molecular correlates of programmed cell death are still subject to disagreements. I predict that, in the next few years, these uncertainties will be resolved largely by experiments driven by theoretical issues such as the importance of bistability, the roles of feedback and feed forward, and robustness in the face of noise.

John Tyson is university distinguished professor of biological sciences at Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA.

FURTHER READING

Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms* (Cambridge Univ. Press, 1996).
Legewie, S., Bluthgen, N. & Herzog, H. *PLoS Comput. Biol.* **2**, e120 (2006).
Locke, J. C. et al. *Mol. Syst. Biol.* doi:10.1038/msb4100018 (2005).
Pomeroy, J. R., Kim, S. Y. & Ferrell, J. E. Jr *Cell* **122**, 565–578 (2005).

For other essays in this series, see <http://nature.com/nature/focus/arts/connections/index.html>



J. KAPUSTA/IMAGES.COM

CONNECTIONS

NEWS & VIEWS

ANIMAL BEHAVIOUR

Planning for breakfast

Sara J. Shettleworth

It is commonly believed that planning for the future is a skill unique to humans. Could other animals, even those as evolutionarily distant as western scrub-jays, share this skill with us?

Can it ever be said that animals plan ahead? Animals do show behaviour that prepares them for the future, but in general that behaviour reflects unlearned or conditioned responses to predictive cues. For example, a swallow flying south or a marmot entering hibernation is reacting to a cue that has foretold the seasons for its ancestors. A hungry rat pressing a lever that provides food in ten seconds, rather than a lever providing food later, does so because rewards are more effective after short than after long delays. Two requirements¹ for genuine future planning are that the behaviour involved should be a novel action, or combination of actions (thus ruling out migrating and hibernating), and that it should be appropriate to a motivational state other than the one the animal is in at that moment (thus ruling out the rat's lever pressing). In their report of two experiments with western scrub-jays (page 919 of this issue²), Raby *et al.* describe the first observations that unambiguously fulfil both requirements.

The scrub-jay (Fig. 1) naturally caches food. In Raby and colleagues' research, jays were first allowed to acquire information about where food would be available in the morning. Then, in a test in the evening, the authors found that the birds behaved as if they were planning for breakfast by caching food items in the place where the food was most likely to be needed. The birds lived in large cages with three compartments (rooms) (see Fig. 1 of the paper on page 919). In the first experiment, each evening they ate powdered pine nuts, food they were unable to cache, in the central room. Then the next morning each bird was confined to one of the end rooms for two hours. In the 'breakfast room', a bird was always fed, whereas in the 'no-breakfast room' no food was given.

The test of planning came after several cycles of this treatment. For the first time, whole pine nuts were provided in the central room in the evening, along with sand-filled trays for caching in the two end rooms. The authors found that the birds cached three times as many pine nuts in the no-breakfast room as in the breakfast room. Importantly, all the data came from this one test: learning how their choices determined the next day's breakfast could not have influenced the jays' behaviour.



R. WILSON/ALAMY

Figure 1 | Food for thought. Raby and colleagues' experiments² with western scrub-jays show that these birds plan for the future by preferentially caching food where it will be needed most.

In the second experiment, the birds learned to expect breakfast in both rooms, peanuts only in one and dog kibble only in the other. On their first opportunity to cache peanuts and dog kibble in the evening, they distributed their caches so as to provide each room with the kind of food it usually lacked.

The results of two recent studies have been proposed as evidence for planning in primates. In one³, monkeys chose between eating four dates and one date. Eating dates makes monkeys thirsty, and the animals received water after a shorter delay if they chose one date. They gradually reversed their natural preference for four dates as if taking account of future thirst. However, this study falls short of the demonstration with the western scrub-jays² because the monkeys underwent repeated trials in which they learnt the consequences of their choices. In the other study⁴, bonobos and orangutans were taught to use a tool to obtain a treat and were then allowed to choose a tool

to take out of the testing room for use when they returned later. Most animals did take the appropriate tool on some trials. But in addition to other problems⁵, individual animals' patterns of success were far from consistent with a true understanding of the task.

Not much more than 100 years ago, interpreting any of these observations as human-like planning would not have been problematic. Indeed, Darwin's⁶ programme for documenting evolutionary continuity between human minds and those of other species encouraged anthropomorphic interpretations of animal behaviour⁷. This attitude was largely replaced, in both experimental psychology⁷ and ethology⁸, by a bias against 'mentalistic' explanations. But recent years have seen a resurgence of attempts to document processes in animals that in humans are accompanied by distinctive conscious states. Besides thinking about the future, examples include awareness of other individuals' states of mind, understanding how

tools work, intentional deception, and empathy. This trend has been impelled by several developments. These include Donald Griffin's⁹ exhortations to ethologists to tackle animal consciousness; legitimization of the study of consciousness in the cognitive sciences generally; and the prospect of understanding the neural and genetic bases of conscious processes using 'animal models'. This last interest partly explains the current resurgence of attempts to study future planning in animals.

In humans, the ability to imagine future events and consciously recollect past ones (episodic memory) is impaired in patients with damage to the region of the brain known as the hippocampus. Furthermore, new imaging studies show that some of the same brain areas are active during both planning and remembering in normal adults¹⁰. Future planning and episodic memory are thus increasingly seen as part of a single human faculty for mental time travel, a faculty that other species have been proposed to lack¹. Of course, as Raby and colleagues² acknowledge, we will never know if a non-verbal animal is actually 'mentally time travelling' anywhere — future or past. Still, it is interesting that birds of the same species that show evidence of future planning also show episodic-

like memory¹¹. In both cases, researchers have cleverly exploited scrub-jays' specialized food caching to tackle questions of general interest in the study of cognitive evolution. But people can plan for, or think back on, much more than breakfast. It remains to be seen whether the abilities that scrub-jays show when caching are similarly applicable in other contexts. ■

Sara J. Shettleworth is in the Departments of Psychology and of Ecology and Evolutionary Biology, University of Toronto, Toronto M5S 3G3, Canada.

e-mail: shettle@psych.utoronto.ca

1. Suddendorf, T. & Busby, J. *Learn. Motiv.* **36**, 110–125 (2005).
2. Raby, C. R., Alexis, D. M., Dickinson, A. & Clayton, N. S. *Nature* **445**, 919–921 (2007).
3. Naqshbandi, M. & Roberts, W. A. *J. Comp. Psychol.* **120**, 345–357 (2006).
4. Mulcahy, N. J. & Call, J. *Science* **312**, 1038–1040 (2006).
5. Suddendorf, T. *Science* **312**, 1006–1007 (2006).
6. Darwin, C. *The Descent of Man and Selection in Relation to Sex* (Murray, London, 1871).
7. Boakes, R. *From Darwin to Behaviourism* (Cambridge Univ. Press, 1984).
8. Kruuk, H. *Niko's Nature* (Oxford Univ. Press, 2003).
9. Griffin, D. R. *The Question of Animal Awareness* (Rockefeller Univ. Press, New York, 1976).
10. Addis, D. R., Wong, A. T. & Schacter, D. L. *Neuropsychologia* **45**, 1363–1377 (2007).
11. Clayton, N. S. & Dickinson, A. *Nature* **395**, 272–274 (1998).

(a molecule that accepts electron pairs from other molecules) to dramatically increase the acidity and reactivity of the H⁺ ion.

So far, so good, but one objective still remained: to accomplish a so-called halocyclization. This is the same type of reaction as described above, but using a halogen atom rather than H⁺. Thousands of exotic halogenated compounds have been isolated from natural sources, many of which show promise as medicinal leads for the treatment of various diseases. Several of these compounds have structures that probably arise from enzymatic reactions resembling the remarkable hopene cyclization, but initiated with a halogen atom⁶. For this reason, synthetic halocyclizations are highly desirable. But there are many inherent problems that must be solved to perform this reaction in the laboratory, which makes the present accomplishment by Ishihara and co-workers⁴ all the more impressive.

The authors required a highly reactive reagent that acts as a halogen source and that gives only one chiral product in a halocyclization reaction (Fig. 1b). They chose to work with iodine, and designed a reagent that is conceptually similar to the chiral H⁺ complex described above. The new reagent comprises a chemical tuxedo that fits around an iodine atom, creating a chiral environment for, and enhancing the reactivity of, that atom. Instead of using a Lewis acid to activate the assembly, the authors used a Lewis base (an electron-pair donor). The resulting 'halocyclase' is capable of converting simple hydrocarbons into iodinated architectures containing carbon rings, with near-enzyme-like control of the enantioselectivity. To explain this phenomenal selectivity, the authors suggest that the starting material and the iodine atom must square up to each other in just one optimal alignment before the reaction can proceed, in much the same way that two people must face each other before they can shake hands.

As with any breakthrough, there is still more work to be done. Although this halocyclization can also be initiated with bromine, the reaction is only enantioselective if iodine is used. The authors get around this problem by demonstrating that iodine atoms in the product can be transformed into other halogens without affecting the all-important chiral form of the product. Another limitation is that the reaction requires one molar equivalent of the chiral promoter — this is inefficient, as, in principle, a smaller quantity of the promoter should suffice. However, the authors have shown that other simple promoters can be used catalytically, if not enantioselectively, thus opening the door for future advances in this area. Finally, more 'halocyclases' are necessary, as many of the halocyclizations reported in this work terminate after the first ring is formed, thereby requiring a second acid-catalysed step to forge the remaining rings.

Before this work, there were no halocyclization methods that approached the exquisite

ORGANIC CHEMISTRY

A tuxedo for iodine atoms

Phil S. Baran and Thomas J. Maimone

Iodine atoms can be fitted with a chemical jacket to control the conversion of simple carbon chains into complex iodine-containing molecules. Previously, such reactions were only possible with enzymes.

The biosynthesis of hopene molecules showcases a spectacular enzyme-catalysed reaction. In one fell swoop, a simple floppy chain of carbon atoms is transformed by a cyclase enzyme into a complex system containing molecular rings, arranged in a well-defined, three-dimensional shape^{1,2}. This cyclization process creates five carbon–carbon bonds and nine stereogenic centres — asymmetric carbon atoms that are especially difficult to prepare in a controlled way. Furthermore, the reaction is enantioselective, meaning that only one of two possible mirror-image (chiral) forms of the molecule is produced. Chemists can only envy this exquisite level of molecular manipulation.

Amazingly, this type of reaction is not unique. It is just one of hundreds of similar naturally occurring transformations used to create the molecules of life. A common thread in these reactions is that a hydrogen atom is transferred from an enzyme to a specific location in the product (Fig. 1a). But other atoms may also be introduced enzymatically at this position, such as oxygen and, more exotically, halogens — chlorine, bromine or iodine.

Although chemists have successfully emulated certain aspects of these cyclization reactions, the incorporation of halogen atoms has been a long-standing challenge. The first bromine-induced cyclization was observed more than 40 years ago³, but only as a side reaction and without any enantioselectivity. On page 900 of this issue, Ishihara and colleagues⁴ report a tremendous advance: the first non-enzymatic, high-yielding, enantioselective cyclization induced by a halogen atom.

In recent years, chemists have edged closer to recreating the power of enzymes in these reactions. Indeed, the challenge of cyclizing substrates using the equivalent of a chiral hydrogen ion (H⁺) — so mimicking the hopene biosynthesis — has already been met by Ishihara and Yamamoto's group⁵, yielding products with a moderate excess of one chiral form over the other. They designed an artificial cyclase that delivers H⁺ to the substrate from one side only. The imitation enzyme was a sort of chemical tuxedo for H⁺ — a molecule that surrounds the ion, simulating the chiral environment found in natural cyclase. The whole assembly was activated by the addition of a Lewis acid

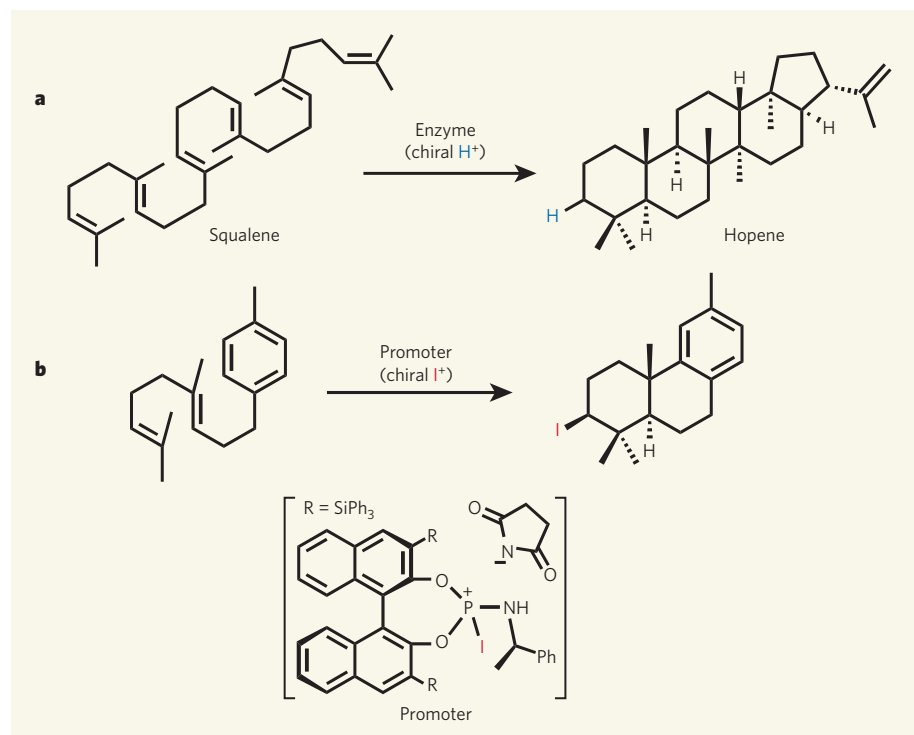


Figure 1 | Enzymatic and synthetic cyclization reactions. **a**, Squalene undergoes a cyclization reaction to form hopene in bacteria, using an enzymatic equivalent of a chiral hydrogen ion (H^+). **b**, Ishihara and colleagues⁴ prepared a chiral source of iodine ions (I^+) that induces a similar halocyclization reaction, so achieving a long-standing goal in organic chemistry. Only one of two possible mirror-image products is made in the reaction.

enantioselectivity of enzymes — it seemed that such fine control could only be achieved with a complex biological catalyst. What is so striking about Ishihara and colleagues' method is that it uses relatively simple reagents. With a chiral jacket for iodine in the closet, the foundation is in place for catalytic versions of this reaction, and for the synthesis of halogenated, naturally occurring compounds. ■

Phil S. Baran and Thomas J. Maimone are in the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla,

California 92037, USA.

e-mail: pbaran@scripps.edu

1. Wendt, K. U., Schulz, G. E., Corey, E. J. & Liu, D. R. *Angew. Chem. Int. Edn* **39**, 2812–2833 (2000).
2. Yoder, R. A. & Johnston, J. N. *Chem. Rev.* **105**, 4730–4756 (2005).
3. van Tamelen, E. E. & Hessler, E. J. *Chem. Commun.* 411–413 (1966).
4. Sakakura, A., Ukai, A. & Ishihara, K. *Nature* **445**, 900–903 (2007).
5. Ishibashi, H., Ishihara, K. & Yamamoto, H. *J. Am. Chem. Soc.* **126**, 11122–11123 (2004).
6. Butler, A. & Carter-Franklin, J. N. *Nat. Prod. Rep.* **21**, 180–188 (2004).

BIOMECHANICS

Fish feeding hardly a drag

Mason N. Dean and Adam P. Summers

Mathematical simulations of prey capture in an aqueous environment, tuned by observational data, have produced a fresh view of the forces generated by suction feeding in fishes.

For humans, suction feeding is a very occasional activity — used to acquire small balls of tapioca from the bottom of a trendy bubble tea, maybe. By contrast, most fishes use suction to obtain all of their food. Typically, a fish targets an individual prey item and swims close, then snaps open its mouth, drawing in a quantity of water along with its lunch. Peter Wainwright and Steven Day¹, writing in the

Journal of the Royal Society Interface, used fluid dynamic modelling and flow data from bluegill sunfish (*Lepomis macrochirus*) to show that the dominant force carrying the prey to its end is not drag from the flowing water, but rather the pressure gradient generated by the rapidly opened mouth.

Several forces, each governed by different equations and generated by the moving fluid,



50 YEARS AGO

Kapitza By A. M. Biew. — This book, written in Germany by a refugee, purports to tell how the U.S.S.R. developed the hydrogen bomb with Kapitza as the principal scientist and with Joffe and Kurtchatov as his principal colleagues... Practically every detail which can be checked is wrong. It is stated that by 1928 Kapitza “had already become in practice the head of the establishment”, that is, the Cavendish Laboratory. This at a time when Rutherford was in his prime! The Royal Society Mond Laboratory, which was built for Kapitza’s work, is referred to as the “Moon Laboratory”. Sensational accounts are given of attempts to lure Kapitza back to the U.S.S.R. in the 1930’s. In fact, he returned most years to see his mother and visit friends... The book states that the Russian atom bomb project started in 1937. While we may be permitted to be sceptical about this, we can at least check the few brief paragraphs about the physics of the project. These appear to be as bogus as the account of Kapitza’s Cambridge period. **J. D. Cockcroft** From *Nature* 23 February 1957.

100 YEARS AGO

The following illustration of Prof. Karl Pearson’s “Random Path” problem may be of interest. Mr. Kipling in his story “The Strange Ride of Morrowbie Jukes” gives the following directions for finding the safe path across a quicksand, which directions are supposed to have been found by the hero of the story in the coat of an earlier victim: — “Four out from crow-clump; three left; nine out; two right; three back; two left; fourteen out; two left; seven out; one left; nine back; two right; six back; four right; seven back.” These numbers were probably taken at random, and it will be noted that seventy five paces are taken, and the final position is only seven paces from the original position. This is a rather curious confirmation of Lord Rayleigh’s solution to the problem. From *Nature* 21 February 1907.

50 & 100 YEARS AGO

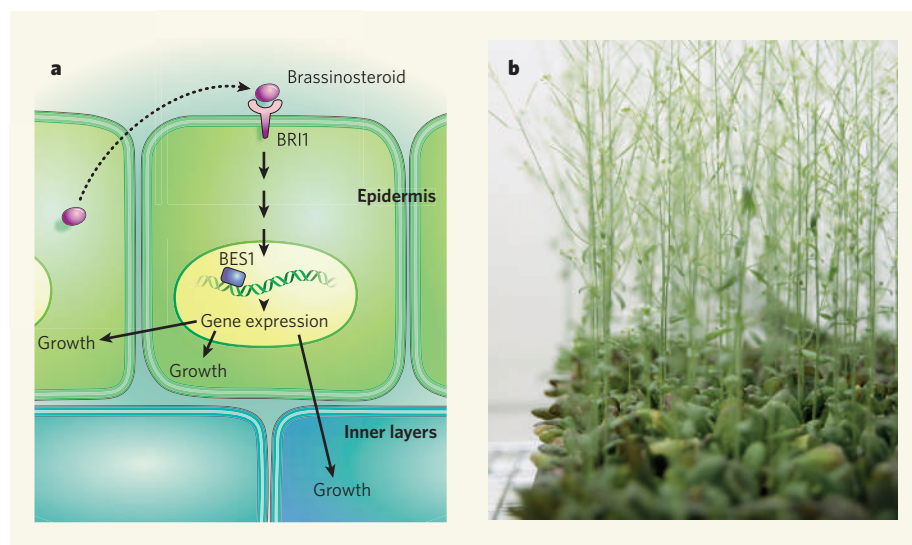


Figure 1 | Control of plant organ growth by signalling in the epidermis. **a**, The plant hormone brassinosteroid binds to its plasma membrane receptor BRI1, resulting in the activation of gene-transcription factors such as BES1, and thus modified gene expression². Savaldi-Goldstein *et al.*³ now show that these signalling events promote growth not only in the epidermal layer, but also in the inner cell layers. Consequently, they suggest that the epidermal layer controls growth throughout the leaves and other shoot-derived organs of flowering plants such as *Arabidopsis thaliana* (**b**).

time, and it has been proposed that the epidermis drives growth 'from without'^{4,5} (Fig. 1a).

Savaldi-Goldstein and colleagues³ set out to study the contribution of the epidermis to shoot growth. They used a mutant of the flowering plant *Arabidopsis thaliana* (Fig. 1b) that lacks both copies of the *CPD* gene. This gene is involved in brassinosteroid biosynthesis and the plant is therefore dwarfed. The authors found that by supplying the epidermal layer of this mutant with an intact version of the *CPD* gene, tagged with a fluorescent 'reporter' molecule, dwarfism could be reversed. When this gene was expressed in the inner tissue layers, it was far less effective in overcoming the growth defect.

Furthermore, dwarfism in plants that had mutations in the *bri1* gene, which encodes the brassinosteroid receptor BRI1, could be reversed by expressing this gene in the epidermal layer only. This confirmed that the receptor-binding ability of brassinosteroids in the epidermis is important for growth. Nonetheless, although growth recovered in these plants, certain aspects of inner-tissue differentiation that also require brassinosteroid signalling — more specifically, the organization of vascular tissues — did not. Moreover, expression of the *bri1* gene only in the vascular layer could not overcome the growth defect.

In an experiment that combined reactivation both of brassinosteroid biosynthesis in the inner tissues and of its receptor-binding ability in the epidermis, the authors confirmed that the inner tissues made only a small contribution to growth. These data convinced Savaldi-Goldstein *et al.* that the brassinosteroid-mediated pathway in the epidermal tissue can drive growth. But they wondered whether the epidermis was also the tissue responsible for

limiting general organ growth under normal conditions. They thus expressed in the various tissue layers a fluorescently tagged enzyme that breaks down brassinosteroids. Consistent with the idea that the epidermis is of major importance for growth, reduced plant stature was particularly obvious when the enzyme was restricted to the epidermis.

These experiments indicate that growth of the epidermal layer in a shoot is crucial for controlling total plant growth. This means that the epidermis must signal to the underlying tissues to keep their growth in step and prevent the plant from being torn apart. The authors therefore investigated how the epidermis might promote the growth of underlying tissues. Plant gene-transcription factors can move through intracellular channels called plasmodesmata. Growth-promoting gene-transcription factors downstream of brassinosteroid signalling have been identified⁶, and Savaldi-Golstein *et al.* showed that tagged versions of these factors could overcome the dwarfism associated with mutations in the *bri1* gene when expressed in the epidermis. One of these transcription factors (BES1) was able to drive even greater growth than was seen in normal plants.

The authors' experiments using fluorescent tags revealed that these transcription factors relay their growth-promoting effect to the inner tissues without moving from the epidermis; thus, the nature of the growth-promoting signal between the epidermis and the inner layers remains unknown. Is the mobile signal chemical in nature? Or is it mechanical, with the epidermis pulling the inner tissues along? One way of addressing this issue is to elucidate the 'growth machinery' regulated by these hormones. Target-gene analysis of transcription factors such as BES1 should identify

which candidates regulate cell expansion in the epidermis and whether these epidermal growth signals influence inner tissues.

Many other questions remain. Do the inner tissues respond only by modulating cell expansion? How does the brassinosteroid growth-control pathway fit in with other mechanisms of growth regulation, for example the regulation of growth repressors such as members of the DELLA family, by various plant growth substances⁷? Is the brassinosteroid response one pathway among several that affect epidermal growth, as is perhaps suggested by the limited overlap in target genes for the different growth-control pathways^{8,9}? Do the various pathways target different layers? And finally, how does an essential role of the epidermis in growth, as observed by Savaldi-Goldstein and colleagues³, relate to genetic analyses in developing flowers, which suggest that the inner tissue layers dictate the identity and thus the development and growth of organs^{10–12}? Perhaps the control of growth involves sequential communication pathways in which the dominance of a layer in one phase of development is followed by dominance of another layer at a later stage. The brassinosteroid pathway dissected by Savaldi-Goldstein *et al.*, together with newly available molecular tools such as fluorescently tagged, functional signal-transduction components, are excellent means to address these questions, and to further our understanding of the integrated control of plant growth.

Ben Scheres is in the Department of Biology, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
e-mail: b.scheres@bio.uu.nl

1. Fujioka, S. & Yokota, T. *Annu. Rev. Plant Biol.* **54**, 137–164 (2003).
2. Li, J. *Curr. Opin. Plant Biol.* **8**, 526–531 (2005).
3. Savaldi-Goldstein, S., Peto, C. & Chory, J. *Nature* **446**, 199–202 (2007).
4. Sachs, J. *Handbuch der Experimentalphysiologie der Pflanzen* (Engelmann, Leipzig, 1865).
5. Van Overbeek, J. & Went, F. W. *Bot. Gaz.* **99**, 22–41 (1937).
6. Yin, Y. *et al. Cell* **120**, 249–259 (2005).
7. Achard, P. *et al. Science* **311**, 91–94 (2006).
8. Nemhauser, J. L. *et al. PLoS Biol.* **2**, e258 (2004).
9. Nemhauser, J. L. *et al. Cell* **126**, 467–475 (2006).
10. Szymkowiak, E. J. & Sussex, I. M. *Plant Cell* **4**, 1089–1100 (1992).
11. Vincent, C. A., Carpenter, R. & Coen, E. S. *Plant J.* **33**, 765–774 (2003).
12. Sieburth, L. E. *et al. Development* **125**, 4303–4312 (1998).

Correction

In the News & Views article "Organic chemistry: A tuxedo for iodine atoms" by Phil S. Baran and Thomas J. Maimone (*Nature* **445**, 826–827; 2006), a couple of errors crept into the accompanying figure. The bond to the top-right hydrogen atom in hopene, which was indicated to be projecting below the plane of the paper, should be pointing upwards; in other words, the hashed bond to this atom should be replaced with a wedge-shaped bond. Similarly, the wedge-shaped bond at the bottom right of the promoter should be replaced with a hashed bond, to indicate that it is projecting below the page, and not above it.

deliver the prey during suction feeding^{2,3}. The relative importance of pressure gradient over drag forces reveals to what degree the predator or prey has control. Drag force is determined by flow speed, which is limited by the predator's mouth shape, but also by the shape and size of the prey item. These two prey qualities are memorialized in the equation for drag force as the coefficient of drag and the area. Prey can reduce the coercion from drag by being more streamlined or smaller. By contrast, the pressure gradient is solely the domain of the predator. Either a smaller mouth or a more rapid expansion of the oral cavity will lead to steeper pressure gradients and therefore greater forces. The prey item has no say in the matter and, if feeding is pressure-dominated, suction capture should not affect the shape of the prey through natural selection.

Wainwright and Day¹ explore the parameters of suction predation by hatching a mathematical model fish and 'feeding' it a variety of prey. The authors' initial abstraction was a fish feeding on a neutrally buoyant, spherical prey of 5 millimetres in diameter and with no escape behaviour. In this simple system they demonstrated that prey within a distance of one-and-a-half mouth widths is completely entrained in the flow into the fish's mouth. Because the prey moves along with the water, there is no relative flow to generate drag forces and capture is determined entirely by the pressure gradient. This approximation is probably accurate for fish feeding on phytoplankton and many of the less mobile aquatic invertebrates.

Not all prey are keen on being lunch, and some try to make a run for it. Surprisingly, when the 5-millimetre sphere is endowed with realistic evasive ability it is still captured, and drag forces make only a small contribution to its demise. Even for zooplankton with a good sensory system and an effective swimming response, death is the inevitable result when a

predator is allowed to approach within a couple of mouth widths. Flow speed drops quickly with distance from the predator, however, so just three or four mouth widths away is beyond the effective range of suction.

In both these examples the flow velocity experienced by the prey is small compared with the velocity of water entering the predator's mouth. This would not be the case for a stubborn little critter clinging to a rock while, for example, a butterfly fish (Fig. 1) tried to suck it off the surface⁴. In this case the water flow over the prey is equal to the flow into the mouth, and drag becomes a more important player — about 30% of the force due to the pressure gradient. For prey items with a good grip on the substrate, drag really does matter but may be irrelevant: some predators may change the game and snatch the prey with a quick nibble.

In the evolutionary arms race between predator and prey, the forces driving prey anatomies and behaviours clearly differ according to the

prey's habitat. In the case of free-swimming prey, the selective pressures should be on the early detection and avoidance of predators. After all, once a fish is close enough to pounce, the prey is gullet-bound regardless of its shape and swimming ability. But for substrate-hugging prey there is a selective pressure to be small enough to fit the roughness of that substrate and streamlined enough to prevent drag forces building up when a butterfly fish comes calling.

Mason N. Dean and Adam P. Summers are in the Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525, USA.

e-mails: mdean@uci.edu; asummars@uci.edu

1. Wainwright, P. C. & Day, S. W. *J. R. Soc. Interface* doi:10.1098/rsif.2006.0197 (2007).
2. Müller, M., Osse, J. W. M. & Verhagen, J. H. G. *J. Theor. Biol.* **95**, 49–79 (1982).
3. Day, S. W., Higham, T. E., Cheer, A. Y. & Wainwright, P. C. *J. Exp. Biol.* **208**, 2661–2671 (2005).
4. Nauwelaerts, S., Wilga, C. D., Sanford, C. & Lauder, G. V. *J. R. Soc. Interface* doi:10.1098/rsif.2006.0180 (2006).

HUMAN GENETICS

Variants in common diseases

Nelson B. Freimer and Chiara Sabatti

Most common diseases arise from interaction between multiple genetic variations and factors such as diet. Studies of such diseases that exploit the rich data on variation in the human genome are just beginning.

The results of the first genome-wide-association (GWA) surveys of common diseases are trickling out. This trickle will soon be a flood of data, much anticipated but challenging to interpret. These initial studies will calibrate our expectations for future investigations, and help to establish the principles for how they are best reported.

On page 881 of this issue, Sladek *et al.*¹ report the results of such a survey of type 2 diabetes*. It is the largest GWA study so far, and tackles a very common disease that is rising in prevalence throughout the world. More than one in three Americans born in 2000 will develop type 2 diabetes, and its rise is particularly rapid in populations that have recently adopted Western lifestyles — hence the efforts to understand the interplay between genetic and environmental risk factors in generating the high frequency of the disease. Sladek *et al.* contribute to these efforts. They demonstrate an unequivocal association between type 2 diabetes and a previously identified² genetic locus (*TCF7L2*), and substantial — but preliminary — evidence for several new loci.

To evaluate GWA studies, we must revise our notion that a discovery in human genetics consists of identifying 'the gene' for a disease. This notion derives from investigations

of rare diseases, which could, in a single study, be associated definitively with mutations in a single gene. Such mutations could be predicted to devastate the function of the proteins encoded by the gene, and thus to cause disease. Yet the 'genetic architecture' of common diseases, such as asthma and depression as well as diabetes, is not built on such obvious deleterious mutations. Rather, it arises from the combined increase in disease risk generated by an unknown number of genetic variants, some of which might not encode proteins, and are thus difficult to identify. In this situation, false positive results are to be expected, and statistically significant results in one study need to be replicated. Even after a gene locus is unequivocally implicated in disease susceptibility, it remains difficult to prove which associated variant is responsible.

For context, it is useful to consider the choices Sladek *et al.* made in terms of the strategy to be used for evaluating genome-wide genetic variation and of the number of individuals to be genotyped. All recent GWA studies assay genome variation using single nucleotide polymorphisms (SNPs); these base substitutions comprise most of the variation in the human genome, and current technology permits economical genotyping of hundreds of thousands of SNPs in a single experiment. Although some studies have successfully used



Figure 1 | Prey's-eye view. A raccoon butterfly fish on the prowl.

*This article and the paper concerned¹ were published online on 11 February 2007.

MUSIC

Let the people sing

Is Joe Public really a flop when it comes to singing? Simone Dalla Bella *et al.* (*J. Acoust. Soc. Am.* **121**, 1182–1189; 2006) went looking for the answer. They persuaded 42 passers-by in a Montreal park to sing *Gens du pays* — an anthem of the Quebec sovereignty movement, commonly sung at festive occasions in the province — on the pretext that it was one of the experimenters' birthday. They then subjected the

recordings, together with those of a further 20 non-musicians and 5 professional singers recorded in the laboratory, to an acoustic analysis.

This quantified errors in timing and in pitch interval (misjudging the jumps between notes) and contour (going the wrong way on the scale). Pitch and timing consistency were assessed by comparing variations in pitch and deviations from the prescribed tempo (*rubato*) in a

repeated phrase in the song's chorus.

Unsurprisingly, the professional participants — among them Gilles Vigneault, the composer and original vocalist of *Gens du pays* — scored better. But they also tended to sing less hurriedly. When, in a second test, the non-musicians were told to slow down, most sang just as accurately as the paid singers.

You could conclude that the ability to sing in tune is a universal human trait. Well, almost — the study uncovered two subjects who, even when singing at the slow tempo, went wildly out of tune.



N. CLEMENTS/TAXI/GETTY

They should perhaps confine their singing to the bathroom.

Richard Webb

SNP sets consisting only of variants within genes^{3,4}, GWA studies aspire to survey all variation in the genome, including non-coding regions. Exhaustive surveys are not yet feasible, but genotyping a subset of the known variable sites may suffice.

Locations in the genome that are separated by a small number of base pairs are often in 'linkage disequilibrium': that is, there is a substantial association between the variants at the two loci. This phenomenon allows us to survey variation across the genome fairly precisely, simply by genotyping a subset of polymorphic loci. GWA studies rely on the assumption that linkage disequilibrium enables one SNP to act as a marker for association to other sequence variants in that region. The GWA studies carried out thus far differ in terms of the number and criteria for selection of the genotyped SNPs: some use SNPs chosen to be evenly physically spaced^{5–7}, whereas others⁸ choose SNPs to maximize the detection of linkage disequilibrium, based on data from the International HapMap Project⁹. Sladek *et al.* used a marker set based on HapMap linkage-disequilibrium data, supplemented by a gene-centric SNP set; their combined marker set (about 400,000 SNPs) provides the highest-resolution survey of genomic variation of any GWA study so far.

In deciding how many individuals to genotype, recent studies follow either a one-stage⁸ or a two-stage^{5–7} design: in the first case, statistical significance is sought within the genotyped sample (and other samples may be used for replication); in the second, SNPs passing a loose significance threshold in the GWA study are genotyped in a follow-up sample, to seek significance. The choice of study design usually rests on estimates of the power to detect effects of a given magnitude that is considered realistic for the trait in question.

Sladek *et al.* adopted a two-stage design: in the first stage, they conducted the GWA study in a total sample of about 1,400 cases and controls (the largest sample in a GWA study so far). In the second stage, they genotyped the SNPs showing evidence of association in the GWA

study in a new sample (about 5,500 total cases and controls) to test for significant associations. The results now published refer only to the follow-up of the most promising SNPs from the first stage: a more comprehensive second stage is still under way.

Sladek *et al.* identify strong associations for three novel loci (one detected in the linkage-disequilibrium-based marker set and two detected in both marker sets). Although more loci may emerge from the complete two-stage analysis, publication of these initial results provides the opportunity for swift replication (or not) by other research groups, using independent samples, as exemplified by the case of *TCF7L2*. Its association with type 2 diabetes was reported last year², and has already been replicated in at least 20 independent studies. In one- or two-stage studies, care must be taken that the specific definition of disease adopted in the follow-up (or replication) samples is comparable to the definition used in the original GWA samples. In this respect it is noteworthy, but of unclear significance, that Sladek *et al.* used more stringent inclusion criteria in the GWA sample than in the follow-up sample.

The identification of a few significant disease associations represents only one outcome of Sladek and colleagues' study. GWA studies should be evaluated primarily from an epidemiological standpoint, focused not just on what new disease-susceptibility genes they propose, but on how they advance our understanding of the composition of genetic risk in the population. Sladek *et al.* take a first step towards such understanding, presenting an evaluation of what proportion of the disease cases can be attributed to variation in the loci they identify as significant in their second-stage analysis. As several additional GWA studies of type 2 diabetes will shortly report their results, we may soon be able to estimate the number — and location in the genome — of the genetic variants that are the main contributors to diabetes susceptibility, at least in some populations.

The results of the first GWA studies may also reveal the degree of genome coverage provided by the chosen SNP panels. Reassuringly, both

of the linkage-disequilibrium-based studies reported so far^{1,8} were able to replicate one known locus. On the other hand, the most significant new locus in both cases was identified by only a single SNP, suggesting that even the dense marker sets employed in these studies provide insufficient coverage for detection of all important loci¹⁰. In certain populations that are of recent origin and that have remained isolated, linkage disequilibrium is more extensive than in the populations used in the GWA studies so far¹¹; the first GWA surveys in such groups will be watched closely for evidence that they permit more complete coverage using comparable marker sets. Similarly, the results of GWA studies, now under way, that are using even larger numbers of SNPs, are much anticipated.

A final observation is that for both of the linkage-disequilibrium-based studies reported so far (for type 2 diabetes¹ and inflammatory bowel disease⁸), the most significant novel associations were to a variant predicted to alter the protein product encoded by a gene (termed a non-synonymous coding SNP), and thus possibly to have a strong functional effect. Furthermore, in both cases the disease is associated with the more common variant at these loci, suggesting that the less common variant may offer protection against developing the disease. That diseases are associated with such common non-synonymous SNPs suggests that these variants may have offered an evolutionary advantage in previous environments. Clearly, one should not generalize from a sample size of two. Nevertheless, these findings underscore how GWA studies may not only deliver 'new' genes¹⁰, but permit advances in our understanding of how human evolution has 'made' the diseases that are common today. ■

Nelson B. Freimer is in the UCLA Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, 695 Charles Young Drive South, Los Angeles, California 90095-1761, USA.

e-mail: nfreimer@mednet.ucla.edu

Chiara Sabatti is in the UCLA Departments of Human Genetics and Statistics, 695 Charles

Young Drive South, Los Angeles, California
90095-7088, USA.
e-mail: csabatti@mednet.ucla.edu

1. Sladek, R. *et al.* *Nature* **445**, 881–885 (2007).
2. Grant, S. F. A. *et al.* *Nature Genet.* **38**, 320–323 (2006).
3. Kubo, M. *et al.* *Nature Genet.* **39**, 212–217 (2007).
4. Hampe, J. *et al.* *Nature Genet.* **39**, 207–211 (2007).

5. Klein, R. J. *et al.* *Science* **308**, 385–389 (2005).
6. Maraganore, D. M. *et al.* *Am. J. Hum. Genet.* **77**, 685–693 (2005).
7. DeWan, A. *et al.* *Science* **314**, 989–992 (2006).
8. Duerr, R. H. *et al.* *Science* **314**, 1461–1463 (2006).
9. The International HapMap Consortium *Nature* **437**, 1299–1320 (2005).
10. Cardon, L. R. *Science* **314**, 1403–1405 (2006).
11. Service, S. K. *et al.* *Nature Genet.* **38**, 556–560 (2006).

GLACIOLOGY

Lubricating lakes

Jack Kohler

More than 150 subglacial lakes have been discovered beneath the Antarctic ice sheet. The four most recent additions, found right at the start of fast flow in a large ice stream, suggest that the lakes influence ice dynamics.

Antarctica is well known as the coldest place on Earth, but it is also surprisingly wet. Whereas the upper part of the ice is cold, geothermal heat and the insulating effect of the overlying ice combine towards its base to bring extensive areas of ice at the bed to its melting point. In the most extreme cases, lakes form beneath the ice sheet. In the past 20 years more than 150 such subglacial lakes have been discovered; Lake Vostok, at 15,690 km² about three-quarters the size of Wales, is the largest and best known¹. As satellites provide better coverage and higher-resolution imagery of the Antarctic ice sheet, more and more such lakes are being found. On page 904 of this issue, Bell *et al.*² present satellite and ground data that reveal four additional lakes in Dronning Maud Land (sometimes referred to in English as Queen Maud Land), East Antarctica. These lakes are remarkable not only for their size, but also for their location, which is just where the flow of the ice sheet begins to accelerate.

Investigations of the base of the Antarctic ice sheet began in the 1960s, when glaciologists used ground-based and airborne radars to map the ice's thickness. Ice is transparent to radar over a wide range of frequencies, so the boundary between ice and bedrock is readily detected, even through several kilometres of ice. Water reflects radar energy more strongly than rock, making it easy to identify those parts of the bed that are wet. Early on, very bright radar reflections delineating a relatively flat boundary were recognized as the distinctive signature of subglacial lakes. The largest subglacial lakes can also be detected using remote sensing: the surface topography above lakes is unusually flat (Fig. 1).

Subglacial lakes were initially thought to be essentially stagnant, with water moving into or out of them at a trickle. But our view of the Antarctic basal drainage system has changed profoundly in light of recent observations. First, satellite radar altimetry data from East Antarctica revealed that the ice surface had dropped suddenly in one area while simulta-

neously rising at another location hundreds of kilometres away^{3,4}. This drainage event, in which large amounts of water had flowed from one subglacial lake to another, occurred over the relatively short period of about 16 months. Laser altimetry data from the West Antarctic ice streams have since exposed a rich variety of lake drainage events, a veritable roiling of the ice surface up and down as water is squeezed from one part of the ice stream to another⁵.

Bell *et al.*² supply the latest evidence for a link between subglacial water and the dynamics of the ice sheet above. The four subglacial

lakes that they have discovered are in the drainage basin of the Recovery Glacier ice stream, which empties into the Filchner Ice Shelf off Antarctica at a longitude of about 30° W. These Recovery lakes are not only located in a part of the continent with relatively few known lakes, but they are also very large; two are exceeded in size only by Lake Vostok. What is most interesting about the Recovery lakes, however, is that they sit at the upstream end of one of the largest of the ice streams that drain the interior ice of the high polar plateau, just where ice velocities start to increase. Upstream of the lakes, the ice sheet moves at a leisurely 2–3 metres per year; downstream of the lakes it suddenly flows at about 50 m yr⁻¹ towards the Filchner Ice Shelf, and eventually into the Weddell Sea.

So how exactly does water underneath an ice sheet affect ice dynamics? As ice flows over a lake, the friction at its base vanishes and its speed increases (see Fig. 1). Furthermore, the freezing of lake water on the underside of the ice sheet adds thermal energy to the basal ice. This prevents the ice from freezing onto the bedrock when it hits the ground again, and also allows faster flow, as warm ice is much softer than cold ice. Nevertheless, for most of the subglacial lakes discovered so far, the surrounding ice apparently buttresses the ice over the lake and prevents faster flow. Bell *et al.* suggest that the location of the Recovery lakes at the onset of a fast-flowing ice stream provides the first evidence that subglacial hydrology could be responsible for the initiation of an ice stream,

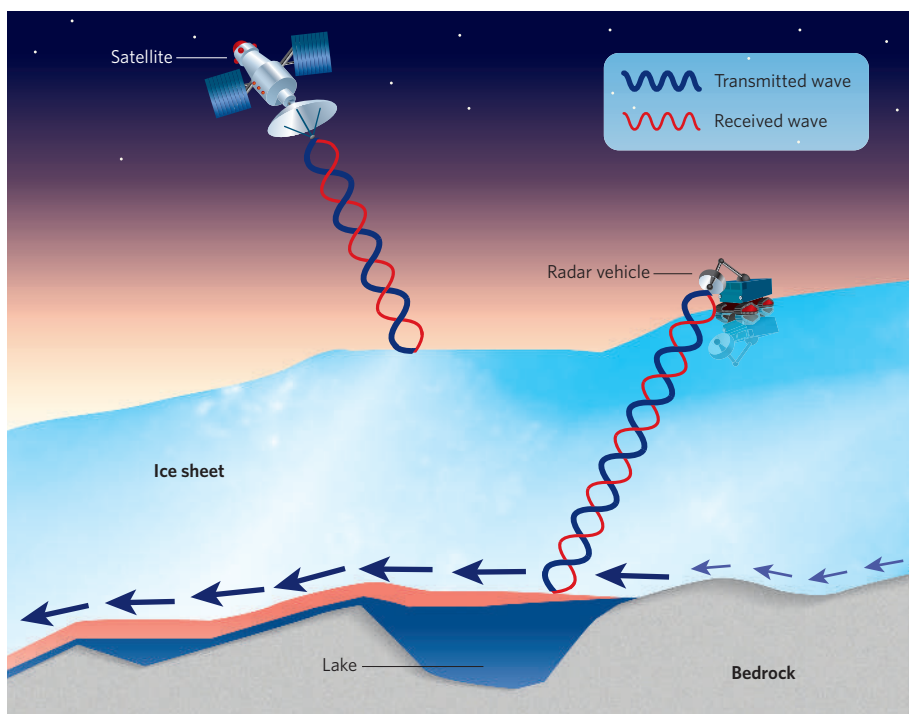


Figure 1 | Slippery slope. The relative strength of reflection of ground-based radar distinguishes between bedrock and water at the base of an ice sheet, and the characteristic flat area in the ice surface above a subglacial lake can be detected by satellite-based surface observations. Using satellite measurements, Bell *et al.*² find four subglacial lakes right at the onset of fast ice flow at the Recovery Glacier ice stream in Dronning Maud Land, Antarctica. They propose that the freezing of lake water to the base of the ice sheet (pink area), together with water spilling over from the lake and scouring the bedrock, could reduce friction and help speed up ice flow towards the sea.

either by changing the ice sheet's basal thermal regime or by introducing water into the path of the ice stream through periodic drainage events.

If this proves to be the case, it would substantially change our understanding of the effect of subglacial lakes on ice-sheet stability. It indicates that basal hydrology needs to be accounted for in numerical models of ice-sheet motion because it has the potential to rapidly induce fast flow. The recent sudden acceleration of ice streams in Greenland^{6,7}, and its resultant effect on the mass balance of the Greenland ice sheet⁸, underscores the importance of understanding the dynamics of ice-sheet response to climatic change.

Might the geographical coincidence of the Recovery lakes and the start of fast flow be just that, a coincidence? So far we have not seen any other large subglacial lakes at the heads of ice streams. But Antarctica remains one of the most poorly mapped parts of our planet — we know more about the surface of the dark side of the Moon than we do about the bedrock or even the surface topography of Antarctica. Antarctica's surface is being mapped at ever higher resolution with remote-sensing satellites such as NASA's ICESat or the planned European Space Agency mission CryoSat-2, but radar imaging of the bedrock is not keeping pace. Proposed missions to put ice-sounding radar instruments on board a satellite, with the goal of mapping the bed in Greenland and Antarctica, have yet to be approved. For now, continued airborne and ground-based geophysical exploration is the only way forward for imaging the ice sheet's base.

During the forthcoming International Polar Year, a number of land traverses and aerial campaigns are planned for Antarctica. In 2008, a Norwegian–US traverse will survey the Recovery lakes area. This was last visited by the US–UK traverse from the South Pole to Dronning Maud Land in 1965–66. Interestingly, that early expedition, which was equipped with one



Figure 2 | Scenes from above the shores of the Recovery lakes. The 1965–66 US–UK South Pole–Queen Maud Land traverse, showing a Sno-Cat tractor and, in the background, a 'rolly' tyre set containing fuel, and the ice-sounding radar antennae. The tractor is just crossing a crevasse at the edge of the Recovery ice stream, near where Bell *et al.*² have now discovered four subglacial lakes.

of the first radars, also crossed over the lakes discovered by Bell and colleagues (Fig. 2). The radar equipment was relatively primitive and used only intermittently, so the lakes were not identified in the field. But a logbook duly noted that Gordon de Q. Robin, then head of the Scott Polar Research Institute in Cambridge, UK, and one of the pioneers of radar glaciology, had suggested the existence of a "...possible melt-layer at the bottom of ice cap".

The implications of this early prediction for ice flow are just starting to be appreciated, as Bell *et al.*² make the link between subglacial lakes and the onset of one of the largest ice streams in Antarctica.

Jack Kohler is at the Norwegian Polar Institute,

Polar Environmental Centre, N-9296 Tromsø, Norway.

e-mail: jack.kohler@npolar.no

1. Kapitsa, A. P., Ridley, J. K., Robin, G. de Q., Siegert, M. J. & Zotikov, I. A. *Nature* **381**, 684–686 (1996).
2. Bell, R. E., Studinger, M., Shuman, C. A., Fahnestock, M. A. & Joughin, I. *Nature* **445**, 904–907 (2007).
3. Clarke, G. K. C. *Nature* **440**, 1000–1001 (2006).
4. Wingham, D. J., Siegert, M. J., Shepherd, A. & Muir, A. S. *Nature* **440**, 1033–1036 (2006).
5. Fricker, H. A., Scambos, T., Bindshadler, R. & Padman, L. *Science* doi:10.1126/science.1136897 (2007).
6. Joughin, I., Abdalati, W. & Fahnestock, M. *Nature* **432**, 608–610 (2004).
7. Luckman, A., Murray, T., de Lange, R. & Hanna, E. *Geophys. Res. Lett.* **33**, L03503 (2006).
8. Zwally, J. *et al.* *Eos Trans. AGU* **87**, Fall Meet. Suppl., Abstr. C14B-03 (2006).

PHYSICAL CHEMISTRY

Oil on troubled waters

David Chandler

The nature of the boundary between water and oil is crucial to many nanometre-scale assembly processes, including protein folding. But until now, what the interface really looks like remained in dispute.

At the boundary between liquid water and water vapour, an interface forms that is marked by an area of lower-than-average density. The same sort of 'depletion layer' also occurs when water comes into contact with a sufficiently large hydrophobic surface — oil, in the most notorious instance, and various other organic molecules. According to theory, this happens because

hydrophobic surfaces provide no opportunity for water molecules to establish their usual hydrogen bonds. Without this adhesive force, the molecules move away from the surface to seek such bonds in the bulk of the liquid.

As lucid as this explanation might seem, experimental confirmation has so far been surprisingly elusive. But at last two independent

groups of researchers — Poynter *et al.*¹, writing in *Physical Review Letters*, and Mezger *et al.*², writing in *Proceedings of the National Academy of Sciences* — provide experimental confirmation of a depletion layer at the interface between water and a hydrophobic surface. And calculations of the molecular dynamics of water confined by two surfaces with nanometre-scale patterns of hydrophobic and hydrophilic regions, described by Giovambattista *et al.*³ in the *Journal of Physical Chemistry C*, provide further grist to the mill.

The prediction that liquid water borders with oil in the same way it does with vapour was first made more than 30 years ago⁴. Since then, the idea has been bolstered by further theoretical analysis⁵, and has helped in understanding the forces that stabilize the assembly of hydrophobic molecules, among them the

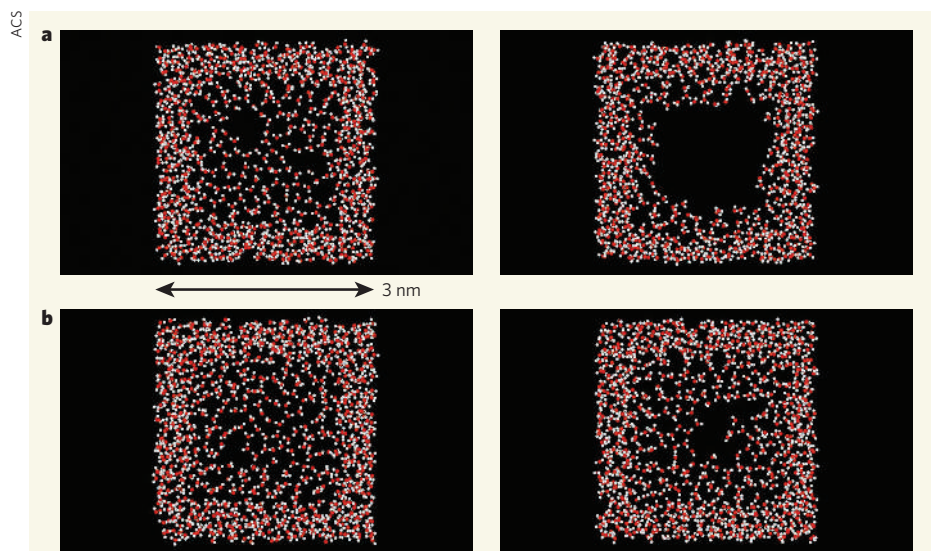


Figure 1 | Depletion in action. In Giovambattista and colleagues' simulations³, water molecules from a large bath of water are confined between two parallel, hydrophobic square plates of side 3 nanometres (to be imagined in front of and behind the plane of the image) separated by 0.6 nanometres. **a**, In this case, one layer of hydrophilic groups is added to ring the outer edges of the otherwise hydrophobic plates. The two images show states between which the system fluctuates; depletion is visible in both, but is most apparent in a large central area in the right-hand image. **b**, If two hydrophilic layers ring the plates, the hydrophobic domain is reduced to a disc less than 2 nanometres in diameter. The degree of density fluctuation is sharply reduced. (Figure adapted from ref. 3.)

building-blocks of proteins, on nanometre and larger length scales^{5,6}. Yet despite these successes, the depletion layer has been such a slippery customer experimentally that its existence has even been called into question⁷.

The reason for this is that, under standard conditions of room temperature and atmospheric pressure, liquid water and vapour very nearly coexist: the free-energy difference per molecule between the two phases is just a fraction of the uncertainty in a molecule's energy that results from ubiquitous thermal fluctuations. An oily surface that is capable of nucleating vapour-like interfaces in water can therefore also nucleate large, metastable vapour bubbles. Such bubbles are transient, non-equilibrium structures, but the density fluctuations that they cause are sufficiently long-lived to confuse observations of the equilibrium interface structure. Furthermore, the interface can wander in space at little free-energy cost. Van der Waals forces between hydrophobic molecules and water limit this translation, but by pulling the interface close to the oily surface, they make the boundary difficult to see.

Poynor *et al.*¹ and Mezger *et al.*² get around the bubble problem by meticulous preparation of their respective hydrophobic organic surfaces. They resolve the interface, which is only a few molecules thick, using synchrotron radiation sources that produce brilliant X-ray beams. The measurements clearly show a layer of depleted average water density that runs along an extended hydrophobic surface — just as the average density of water within a few tenths of a nanometre of its intrinsic liquid–vapour interface is depleted relative to that of the bulk liquid.

Large density fluctuations such as those that

make X-ray measurements difficult are intrinsic to hydrophobic forces that guide molecular assembly. Bulk liquids are almost incompressible and strongly resist density fluctuations needed to accommodate solvation — the reorganization of solvent molecules that occurs when a solute molecule is introduced. The associated high free-energy cost can be balanced if there are very strong attractive forces between solvent and solute. Hydrophobic species have no such forces to compete with the hydrogen bonding between water molecules, so their solvation is generally costly.

Nevertheless, if a large enough cluster of a hydrophobic species does manage to become hydrated, a hydrophobic surface of low curvature may extend over a nanometre or more. In this case, the nucleation of a soft, vapour-like interface with the liquid is energetically favoured. Unlike in bulk liquids or at solid surfaces, the resulting fluid interfaces can be deformed by forces as weak as those produced by typical thermal fluctuations. The mechanism of hydrophobic assembly so important to protein structuring is thus apparent: the reorganization of the solvent required to accommodate small, separated hydrophobic species is more costly than is the accommodation of the clustered species, provided the cluster is nanometre-scale or larger.

In measuring a depleted water density where water meets an extended hydrophobic surface, the new experiments^{1,2} have verified the existence of interfacial fluctuations indirectly. If the average density is, say, half that of bulk water, this can only mean that the instantaneous density is as often at the bulk value as it is at the vapour value. Further experimental work is required to see the fluctuations explicitly.

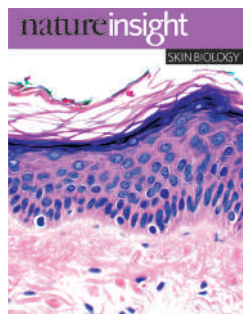
In computer simulations of liquid water with large hydrophobic solutes⁸, fluctuations were noted as long ago as 1995. They were further inferred from the observation several years later of huge fluctuations in forces required to trap a film of water between hydrophobic and hydrophilic solid surfaces⁹. A simulation of a hydrated carbon nanotube provides another case in point¹⁰: in this example slight adjustment of parameters proves sufficient to move the system from a regime in which the tube is mostly filled with water, to another in which the tube is mostly empty. The dynamics of the transformation involve collective motion of many water molecules, suggesting the possibility of nanometre-scale steam engines that are perhaps already harnessed in some biomolecular motors.

Giovambattista and colleagues³ add illustrations of an aspect related to these nanometre-scale manifestations of fluctuations and liquid–vapour coexistence. Altering the hydrophobic surface, for example by the inclusion of hydrophilic units, should bring significant changes to the interface. Natural hydrophobic systems such as proteins generally have heterogeneity of this sort: it is their 'amphiphilic' character that allows them to be solvated at reasonable cost. Giovambattista *et al.* perform simulations of water confined between patterned hydrophilic and hydrophobic surfaces, and record collective fluctuations evocative of dynamic equilibrium between liquid and vapour (Fig. 1). Hydrophilic heterogeneities tend to pin the soft fluid interface to the otherwise oily surface, and a large enough hydrophilic region removes it altogether. Increasing pressure moves water away from phase coexistence, and the authors show that enough pressure removes the interface, and the mean depletion associated with it.

Such computational work throws the gauntlet back to the experimentalists, to see whether they can observe such behaviour in the laboratory. Much could ride on the answer, as knowing the timescales and energetics of interface fluctuations, as well as their sensitivity to hydrophilic inhomogeneities, could be central to understanding kinetics of self-assembly. One only hopes that this time we will not have to wait another 30 years for experimental validation. ■ David Chandler is in the Department of Chemistry, University of California, Berkeley, California 94720, USA.

e-mail: chandler@cchem.berkeley.edu

- Poynor, A. *et al.* *Phys. Rev. Lett.* **97**, 266101 (2006).
- Mezger, M. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 18401–18404 (2006).
- Giovambattista, N., DeBenedetti, P. G. & Rossky, P. J. *J. Phys. Chem. C* **111**, 1323–1332 (2007).
- Stillinger, F. H. *J. Solut. Chem.* **2**, 141–158 (1973).
- Lum, K., Chandler, D. & Weeks, J. D. *J. Phys. Chem. B* **103**, 4570–4577 (1999).
- Chandler, D. *Nature* **437**, 640–647 (2005).
- Ball, P. *Nature* **423**, 25–26 (2003).
- Wallqvist, A. & Berne, B. J. *J. Phys. Chem.* **99**, 2893–2899 (1995).
- Zhang, X., Zhu, Y. & Granick, S. *Science* **295**, 663–666 (2002).
- Hummer, G., Rasaiah, J. C. & Noworyta, J. P. *Nature* **414**, 188–190 (2001).

**Cover illustration**

Haematoxylin and eosin stain of normal human skin showing morphology of dermis and epidermis. (Image courtesy of S. R. Granter, Brigham and Women's Hospital, Massachusetts.)

Editor, *Nature*

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Lesley Anson

Production Editor

Sarah Archibald

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Sponsorship

Claudia Banks
Gerard Preston
Emma Green

Production

Susan Gray

Marketing

Katy Dunningham

Editorial Assistant

Laura Shaw

SKIN BIOLOGY

Not only is skin the largest organ of the body, it is also one of the most multifaceted. It senses our surroundings and forms a protective barrier against many different environmental insults, including ultraviolet radiation and pathogens.

As our interface with the outside world, skin — through its ability to sense touch and temperature — can be the source of much delight. But where there is pleasure, there is pain. For example, skin senses extreme temperatures and, as a result, its nerve endings perceive pain. Its intricate structure and biological functions are associated with various afflictions, including developmental defects, autoimmune disorders, allergies and cancer. Skin (and the hairs that grow from it) also forms a central aspect of our appearance that can bring about much anxiety and prejudice.

Although the development of skin and hair is a sophisticated, finely tuned process, it is now possible to generate artificial skin as a model system for research and for the treatment of burns patients. Constant progress is being made to improve this approach to tissue engineering.

This Insight brings together review articles that explore the complex molecular processes that underlie skin and hair development and pigmentation, the mechanisms of sensory transduction that originate in the skin, and the cellular and molecular changes responsible for skin cancer and psoriasis. It also includes discussion about how this knowledge could guide researchers to develop better therapies for diseases that affect the skin.

We are pleased to acknowledge the financial support of Johnson & Johnson Consumer Companies, Inc. in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content and peer review.

**Barbara Marte, Joshua Finkelstein and Lesley Anson,
Senior Editors**

REVIEWS

834 **Scratching the surface of skin development**

E. Fuchs

843 **Melanocyte biology and skin pigmentation**

J. Y. Lin & D. E. Fisher

851 **Melanoma biology and new targeted therapy**

V. Gray-Schopfer, C. Wellbrock & R. Marais

858 **Mechanisms of sensory transduction in the skin**

E. A. Lumpkin & M. J. Caterina

866 **Pathogenesis and therapy of psoriasis**

M. A. Lowes, A. M. Bowcock & J. G. Krueger

874 **Progress and opportunities for tissue-engineered skin**

S. MacNeil

nature
insight

Scratching the surface of skin development

Elaine Fuchs¹

The epidermis and its appendages develop from a single layer of multipotent embryonic progenitor keratinocytes. Embryonic stem cells receive cues from their environment that instruct them to commit to a particular differentiation programme and generate a stratified epidermis, hair follicles or sebaceous glands. Exciting recent developments have focused on how adult skin epithelia maintain populations of stem cells for use in the natural cycles of hair follicle regeneration and for re-epithelialization in response to wounding.

At the surface of our body, the skin epidermis and its appendages provide a protective barrier that keeps microbes out and essential body fluids in. It receives daily assaults, including harmful ultraviolet radiation from the Sun, and scratches and wounds. It confronts these attacks by undergoing continual self-renewal to repair damaged tissue and replace old cells. To do this, it depends on stem cells, which reside in the adult hair follicle, sebaceous gland and epidermis for the purpose of maintaining tissue homeostasis, regenerating hair and repairing the epidermis after injury. Where do these adult stem cells come from and when during embryonic development are they determined? How do they specify on demand one differentiation programme versus another, and what governs these lineages? How do stem cells in these tissues know how many cells to replenish and when? And, finally, how do the skin epithelial cells communicate with the immune system to prevent infection? Here I place particular emphasis on recent studies to review what is known about these fascinating features of mammalian skin epithelium from embryogenesis to adult. I also speculate on how dysregulation of the normal process of wound healing in the skin epidermis can lead to skin disorders, including cancers and chronic inflammation. The review concentrates heavily on mouse skin, given the accelerated pace of scientific discoveries made possible by mouse genetics.

Embryonic origins of skin epithelium

After gastrulation, the embryo surface emerges as a single layer of neuroectoderm, which will ultimately specify the nervous system and skin epithelium. At the crossroads of this decision is Wnt signalling, which blocks the ability of ectoderm to respond to fibroblast growth factors (FGFs). In the absence of FGF signalling the cells express bone morphogenetic proteins (BMPs), and become fated to develop into epidermis. Conversely, the acquisition of neural fate arises when, in the absence of a Wnt signal, the ectoderm is able to receive and translate activating cues by FGFs, which then attenuate BMP signalling through inhibitory cues¹ (Fig. 1a). The embryonic epidermis that results from this process consists of a single layer of multipotent epithelial cells. It is covered by a transient protective layer of tightly connected squamous endodermis-like cells, known as a periderm, which are shed once the epidermis has stratified and differentiated².

Hair formation

Mesenchymal cells from the dorsal backskin dermis are derived from dermomyotome, in which Wnt signalling specifies their fate³. As these mesenchymal cells populate the skin, their interactions with the overlying epithelium induce the formation of hair placodes, which appear as small epidermal invaginations into the underlying dermis. Pioneering studies on mesenchymal–epithelial tissue recombination in chicks

and mice revealed that early cues from the mesenchyme determine the positioning of placodes and specify their commitment^{4,5} (Fig. 1). Key components of these mesenchymal–bud-promoting signals include FGFs^{6,7} and, subsequently, BMP-inhibitory factors^{8,9}. In the presence of excess BMPs, or the absence of either the BMP-inhibitor noggin or the FGF10/FGF7 receptor FGFR3b, follicle density is reduced^{7,10,11}.

In conjunction with these early dermal signals, ectodermal Wnt signals instruct the epidermal cells to grow downwards to form a hair bud or placode^{12–14}. Among the early genes expressed in the placode are sonic hedgehog (*Shh*), which has a crucial role in organizing the dermal cells into a condensate, or dermal papilla, which thereafter becomes a permanent associate of the hair follicle^{15–17}. Once formed, the dermal papilla signals to the placode to grow downwards to form the hair follicle, which matures through additional reciprocal mesenchymal–epithelial crosstalk⁴.

Increasing evidence places Wnt signalling at the helm of placode formation. Although the single layered epithelium uniformly expresses a number of Wnts, including Wnt10b, the placode and dermal condensate seem to respond most robustly, as judged by the presence of nuclear β -catenin, the outcome of a Wnt signal^{13,14,18,19}. Embryonic placodes and dermal papillae also exhibit activation of the Wnt reporter gene *TOPGAL*. This drives β -galactosidase expression under the control of an enhancer composed of multiple recognition sites for the LEF1/TCF family of DNA-binding proteins, which function concomitantly with Wnt/ β -catenin signalling^{13,19,20}. LEF1/TCF proteins can act as transcriptional repressors, but upon association with β -catenin act as a functional bipartite transcriptional activator. On the basis of timing and morphology, Wnt activity in the epithelium seems to precede that in the dermal condensates of developing hair follicles.

The TCF family member TCF3 is expressed early in embryogenesis when the skin epithelium exists as a single layer of unspecified progenitor cells²¹. Placode initiation and Wnt reporter activity are accompanied by a reduction in TCF3 and upregulation of LEF1. LEF1 concentrates in developing hair placodes and the dermal papilla. TCF3 seems to maintain epithelial progenitor status: when it is sustained transgenically in these progenitors, all three differentiation lineages (epidermis, hair follicle and sebaceous gland) are repressed²¹. Conversely, LEF1 is important for follicle morphogenesis, and when *Leff1* is knocked out a reduction in hair follicles is observed²².

Mice expressing a constitutively stabilized form of β -catenin in the epidermis develop excess hair follicles^{12,23–25}. When β -catenin is conditionally targeted or when the Wnt inhibitor dickkopf-1 (DKK-1) is expressed ectopically, epidermal appendage formation is impaired^{26,27}. On the basis of these findings, Wnt signalling seems to be an integral part of the early epidermal–dermal discourse that specifies hair patterning and morphogenesis.

¹Howard Hughes Medical Institute, Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, 1230 York Avenue, Box 300, New York, New York 10021, USA.

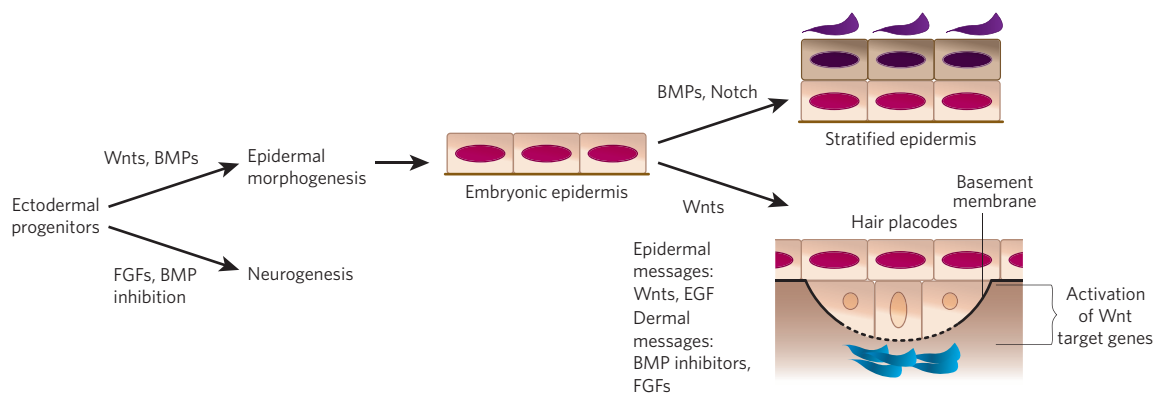


Figure 1 | Early signalling steps in specification of embryonic skin.

a, In the absence of Wnt signalling, ectodermal progenitors respond to FGFs, downregulate BMP signalling and progress towards neurogenesis¹. Wnt signalling blocks the ability of early ectodermal progenitor cells to respond to FGFs, allowing them to respond to BMP signalling and adopt an epidermal fate. As development progresses, single-layered embryonic epidermis expresses Wnts. Some cells fail to respond to Wnts,

and these become fated to become epidermal cells through BMP, EGF and Notch signalling. The cells that do respond to Wnt signalling also receive underlying FGF and BMP inhibitory signals from the mesenchyme and, together, these signals instruct the cells to make an appendage^{4,6–10}. Collectively, the inhibition of BMP inhibitory signals and Wnt activating signals produce the hair placode^{9,12,13}. Additional dermal messages from below further instruct the placodes to make the follicle.

In apparent contrast to the epidermal/neural fate switch, activation of Wnt signalling and inhibition of BMP signalling seem to be on the same side of the epidermal/hair placode switch. Thus, the β -catenin/LEF1-responsive, *TOPGAL*-positive hair placodes lack nuclear phosphorylated (activated) SMAD1, the outcome of BMP signalling¹¹. Moreover, in mice that lack noggin, ectodermal LEF1 is no longer expressed and *TOPGAL* Wnt reporter activity is diminished²⁸. Conversely, LEF1 expression is maintained upon conditional ablation of *Bmpr1a*, which encodes the key BMP receptor for the epidermis and its appendages^{29,30}. If inhibition of BMP signalling promotes LEF1 accumulation, this could explain in part how BMP inhibitory and Wnt-promoting signals converge in placode formation.

Hair follicle maturation

Once the placode has formed, downstream signalling events drive the downgrowth and maturation of the hair follicle. A myriad of changes take place during follicle morphogenesis, as exemplified by the differences between the transcriptional profile of placode cells and their epidermal counterparts³¹. The purified placode cells are positive for *TOPGAL*, and the large number of Wnt pathway and Wnt target genes comprising the placode signature is reflective of the programme of gene expression that occurs as a downstream consequence of Wnt signalling.

One of the key regulators downstream of Wnt/noggin is SHH. Embryonic skin deficient in SHH still expresses *TOPGAL* and LEF1 and shows placode formation^{15,16,31}. However, in its absence the overall number of placodes is diminished and they do not progress past the hair germ stage, although a small number of terminally differentiated hair cells can still be detected. Microarray analyses of developing wild-type hair placodes have confirmed the elevated expression of SHH effectors such as GLI-Kruppel family member 1 (GLI1) and patched homologue 2 (PTC2)³¹. In addition, transcriptional profiling has uncovered genes encoding a possible SHH-regulatory circuit involving the transcription factors FOXC1, TBX1, DACH1, SIX1 and LHX2, whose family members have been shown to mediate eye development³¹. Loss-of-function mutations in at least one of these downstream effectors, *Lhx2*, result in a sparsity of follicles and a failure to maintain follicle stem-cell behaviour in mice. Thus, the functions of SHH signalling seem to extend beyond merely governing proliferation and follicle downgrowth.

It is curious that with the exception of β -catenin, whose loss affects all follicle formation, the loss of other crucial inductive factors, such as SHH, noggin, LEF1 and LHX2, diminishes rather than abolishes hair germ and/or follicle formation. One clue to understanding these differences is provided by the fact that mice defective in ectodysplasin-A receptor (EDAR)/EDA signalling lack guard hair follicles, whereas mice lacking

LEF1 have guard hairs but lack other types of hair follicle^{11,32,33}. Underlying these variations seems to be the different ways in which BMP inhibition can be achieved in order to permit transmission of the follicle Wnt signal. Thus, guard hairs rely on EDAR/EDA/nuclear factor- κ B (NF- κ B) signalling, which leads to activation of the BMP inhibitor CTGF (connective tissue growth factor)¹¹; the other follicle types seem to be more dependent on noggin for BMP inhibition and LEF1 activation²².

Whereas different modes of BMP inhibition affect the types of follicle that are specified, active BMP signalling seems to affect follicle density once specification has occurred. Soon after the placode forms, BMP4 levels are increased, even though it is the ectoderm, and not the placode, that expresses nuclear phospho-SMAD1 — that is, recognition that a cell is responding to a BMP — at this time³¹. The importance of this molecular idiosyncrasy seems to be rooted in negative feedback regulation, as suggested by the fact that nascent follicles always emerge at the apexes of a hexagon at the centre of which is an existing follicle from the prior wave of follicle morphogenesis. Such spacing places follicles at the maximum distance from one another, and is suggestive of an inhibitory signal transmitted by the placode to restrict its neighbours from staking their claims too closely. Placodes also express Wnt inhibitory factor 1 (WIF1), suggesting that the negative feedback circuitry might be complex. The dermis also influences follicle size and density, and the recent molecular differences found in dermal cells from different body sites seem likely to expedite progress in deciphering these anatomical distinctions in epithelial–mesenchymal interactions³⁴.

The epidermis

At first glance, it might seem as though the epidermis is simply a default option in the crossfire of signalling pathways directed at making an appendage. However, it is important to consider that BMP signalling is active in the interfollicular epidermis (IFE), and, as such, is as much an epidermis-promoting signal as it is a follicle-inhibitory signal^{11,29,30} (Fig. 1). In addition, studies on chicks indicate that epidermal growth factor receptor (EGFR) signalling occurs in the epidermal cells at the boundary of each feather placode. Moreover, elevated EGF promotes the proliferation and expansion of interbud epidermal cells, whereas inhibition of EGFR signalling results in increased acquisition of feather bud fates³⁵. Whether EGFR signalling has a similar role in inhibiting the formation of hair follicles is less clear, although elevated EGFR signalling in mammalian skin epithelium is associated with enhanced epidermal proliferation and hair loss³⁶. Thus, when taken together, follicle density seems to be governed by competition between placode-stimulating (epidermis-restricting) and placode-restricting (epidermis-stimulating) signals.

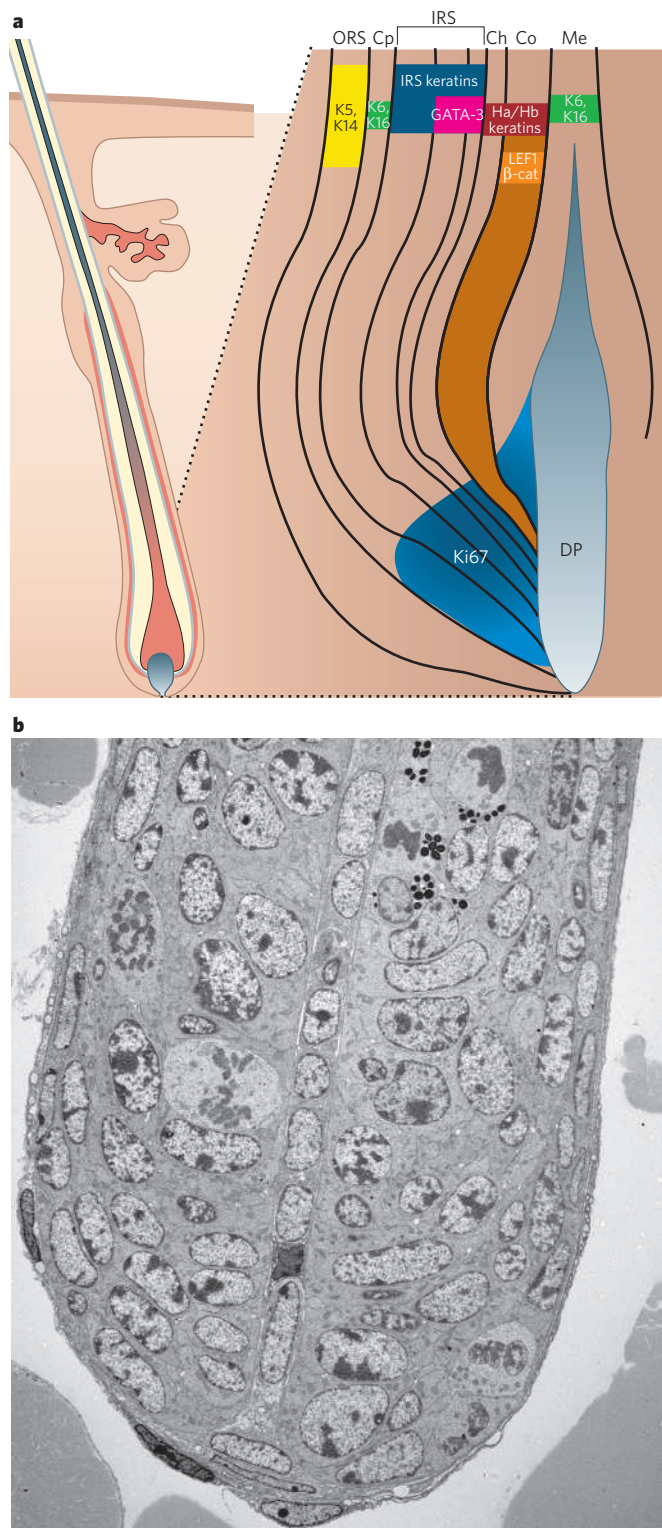


Figure 2 | Rapidly proliferating matrix-cell progenitors are spatially organized in the hair bulb to respond to distinct differentiation-specific cues. **a**, Schematic of the hair bulb. Rapidly proliferating matrix cells at the base (blue) give rise to seven differentiation-specific lineages. The various genetic markers for these lineages are shown. Matrix cells express the proliferative marker Ki67. As these cells differentiate, they express keratins differentially. Ch, hair shaft cuticle; Co, cortex; Cp, companion layer; DP, dermal papilla; Me, medulla^{29,46}. **b**, Ultrathin section of hair bulb. The single vertical layer of core cells corresponds to the dermal papilla. Note the onion-skin-like organization of matrix cells and their differentiating progeny at the base of the follicle. (Image courtesy of H. A. Pasolli, The Rockefeller University, New York.)

Hair follicle maturation and lineage differentiation

As an embryonic hair follicle grows downwards, its leading front (matrix) is highly proliferative and, through its sustained contact with the dermal papilla, adopts a programme of gene expression that is distinct from that at the root. The outer root sheath (ORS) maintains contact with the basement membrane while the inner layers begin to differentiate to form the inner root sheath (IRS), which will serve as the channel for the protruding hair (Fig. 2a). At about embryonic day (E) 18.5 in mice, the earliest vestiges of the central core of hair shaft cells emerge coincident with the appearance of the sebaceous gland precursor cells in the upper part of the hair follicle. Maturation continues for another 7–8 days, as a companion layer separates the ORS and IRS, and the IRS and hair shaft each develop further into three concentric cell layers (Fig. 2a). The discrete stages of hair follicle development are readily distinguished by their morphological and biochemical differences, and have been described elsewhere³⁷.

Within each mature hair follicle the seven concentric rings of terminally differentiating cells are derived from matrix cells. Each ring has a distinctive ultrastructure (Fig. 2b). Closer inspection of the hair bulb provides some insight into how these rings might form, and what driving forces might orchestrate their differentiation. At the height of the growth phase (full anagen) of the follicle the dermal papilla exists as a thin strand of cells, separated from the hair bulb surrounding it by a basement membrane. Some mitotic matrix cells seem to orient their spindle plane at an angle relative to the basement membrane, suggesting that distinct differentiation lineages might be generated in an onion-skin-like fashion. Lineage tracing experiments provide support for this model^{17,38,39}, which is similar to that postulated for the epidermis⁴⁰ (discussed below).

Hair follicle lineages

During the years, researchers have uncovered molecular differences that are diagnostic for the lineages in the hair follicle. The most prominent structural proteins of these lineages are keratins (Fig. 2a), which are differentially expressed in specific patterns that distinguish most, if not all, of the concentric layers within the hair follicle⁴¹. Efforts to understand how these keratin genes are differentially regulated led to the discovery that the hair shaft *Ha/Hb* keratin genes are genuine Wnt target genes, and that Wnt signalling has a crucial role in the differentiation of matrix cells along a hair shaft lineage¹⁹. Constitutively active β -catenin in humans and mice leads to pilomatricomas — hair tumours composed exclusively of the hair lineage^{12,42}. Additional hair shaft transcriptional regulatory proteins of importance are HOXC13 and FOXN1, both of which cause hair defects when mutated^{43,44}.

By contrast, mutations in the genes encoding GATA-binding factor (*Gata-3*) and CCAAT displacement protein (*Cdp*) yield primary IRS defects, which lead to alterations in the shaft^{45,46} (Fig. 2a). GATA-3 and CDP are both able to recruit histone deacetylases (HDACs) and transcriptionally repress their target genes. On the basis of studies in other tissues, candidate genes for CDP-mediated repression in the IRS include *c-myc*, *N-cam* and *p21*. GATA-3 represses its own expression in the IRS, but its other targets remain unknown. The transcriptional repressor BLIMP1 also concentrates in differentiating IRS cells, and like CDP, one of BLIMP1's target genes is *c-myc*^{47–49}. If BLIMP1 associates with a histone arginine methyl transferase as it does in germ cells^{47,48} then it could add an epigenetic mark to the chromatin it represses in the skin. Together, these observations suggest that epigenetic remodelling of chromatin might have a crucial role in the IRS lineage. Such hints merit further investigation in the future.

Although at first glance it might seem that IRS differentiation is governed by repression and relief of repression, positive-acting signalling pathways are also involved. One of these is Notch, a transmembrane receptor that, when activated by its ligand, can be cleaved by γ -secretase to generate a nuclear intracellular cofactor (NICD) for the transcriptional repressor protein RBP-J⁵⁰. Although Notch signalling is not required for follicle specification, IRS cells of γ -secretase-deficient follicles fail to maintain their fates; perhaps as a consequence, ORS cells aberrantly activate epidermal differentiation, hair differentiation is compromised and follicles develop into epidermal cysts^{51,52}.

BMP receptor signalling is also involved and is essential for both IRS and hair shaft differentiation. In the absence of BMPRI1A, masses of undifferentiated, highly proliferative matrix-like cells accumulate that rapidly progress to tumours⁵³. The fact that epidermal cysts arise from Notch signalling defects whereas matrix-like cysts result from BMP signalling defects underscores distinct differences in the ways these pathways affect hair follicle lineages. Sifting through the underlying mechanisms involved and the myriad of factors that affect fate specification and differentiation of follicle matrix cells is one of the main challenges currently facing skin biologists.

The hair cycle

The hair follicle is one of the few organs of the body that undergoes cyclic bouts of degeneration and regeneration throughout life (Fig. 3). Matrix cells are sometimes referred to as transit-amplifying cells

because they only survive through the growth (anagen) phase of the cycle. After the first 2 weeks of postnatal mouse life, the initial supply of matrix cells declines, hair shaft and IRS differentiation slow, and the follicle enters a destructive phase known as catagen⁵⁴. During this phase, which in mice lasts for 3–4 days, apoptosis reduces the follicle to an epithelial strand. This drags the dermal papilla upward to rest just below the permanent, non-cycling upper follicle.

A number of molecular regulators of the anagen–catagen transition have been identified, although how they work together to promote catagen or terminate anagen and how they spare progenitor cells needed for anagen re-entry is not yet understood. Recently, keratin 17 was found to be a regulator of this process. It seems to regulate anagen by interacting with and sequestering a death adaptor protein required for tumour-necrosis factor receptor 1 (TNFR1)-dependent signalling and induction of apoptosis to initiate the catagen process⁵⁵. In addition, genetic

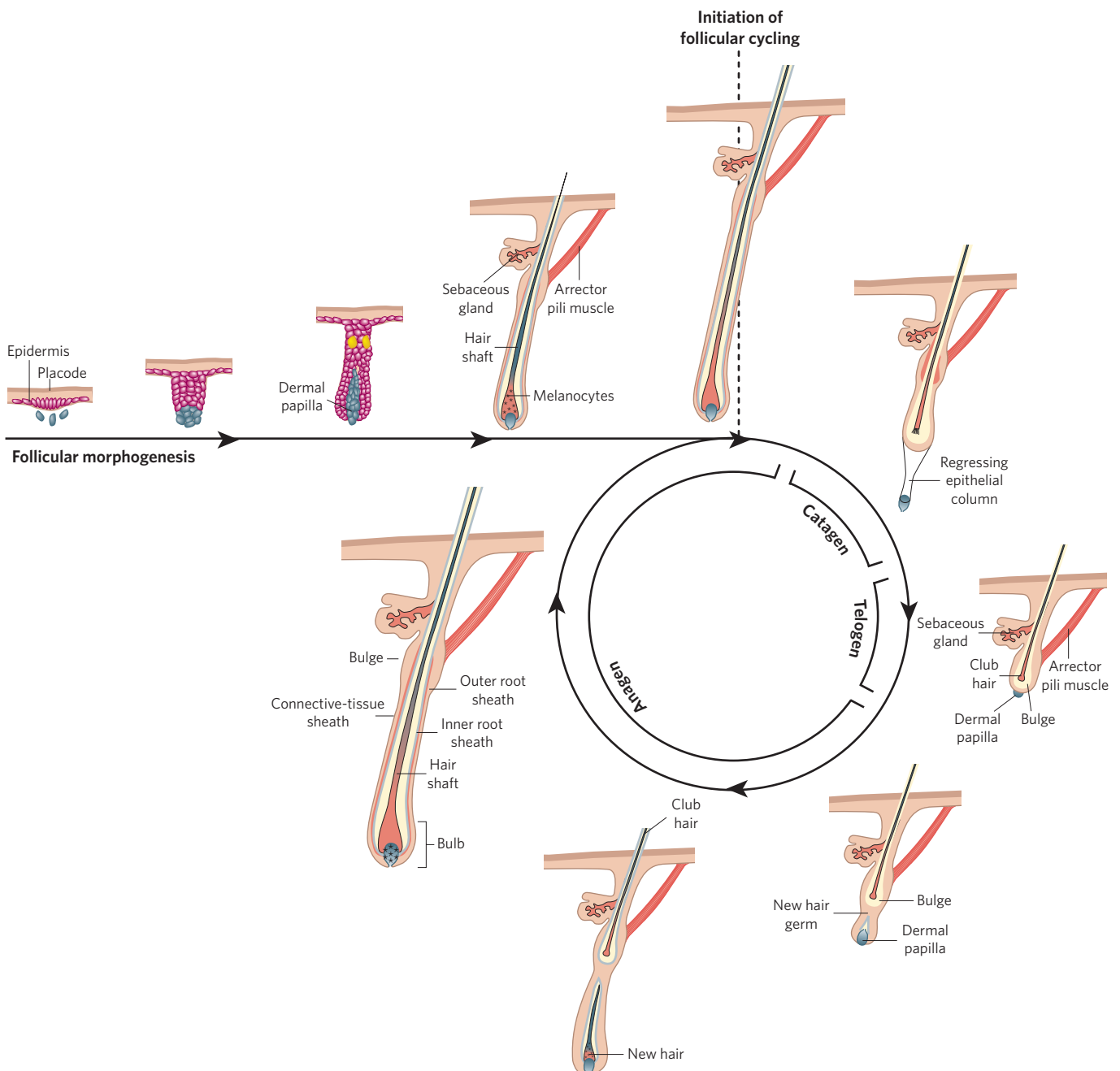


Figure 3 | The hair cycle. The stages of the hair cycle are depicted, starting from the first postnatal anagen, when the hair shaft is growing and protruding through the skin surface. Follicles progress synchronously to the destructive (catagen) phase, during which the lower two-thirds of the

follicle undergo apoptosis and regress. The dermal papilla is brought to rest below the bulge-stem-cell compartment, and after the resting (telogen) phase, a critical threshold of activating factors is reached and the stem cells become activated to regrow the hair¹⁰⁰.

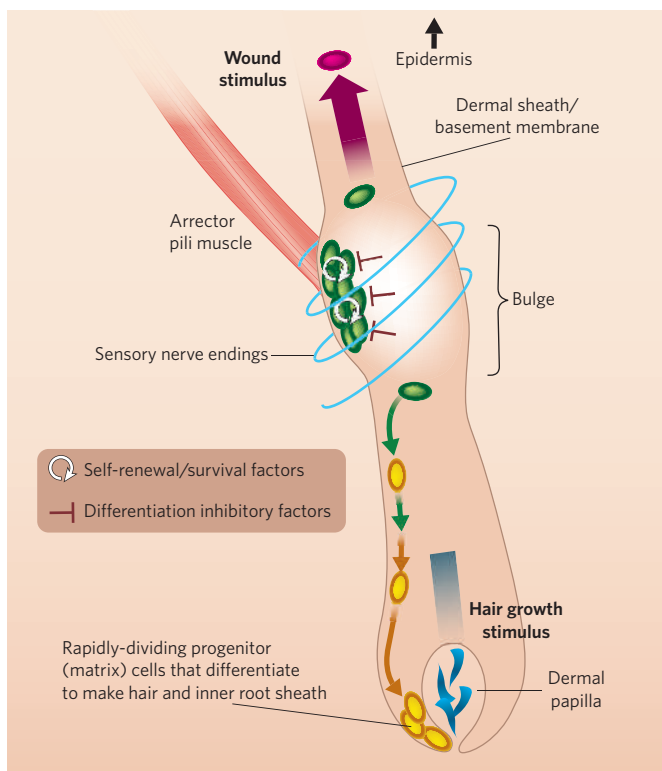


Figure 4 | Diagram of the follicle stem-cell niche. The bulge contains infrequently cycling, label-retaining cells, which include multipotent stem cells (green) that can generate the new hair follicle during cycling and repair the epidermis on injury. The bulge is in a specialized niche, surrounded by other cell types, which together provide cues that maintain these cells in an undifferentiated, quiescent state. For stem cells to be activated, the niche environment must change.

evidence has implicated a transcription complex involving the vitamin D-retinoid X receptor- α transcription factor and an interacting repressor protein, *hairless*. When defective, this complex impairs the catagen-related defects in the epithelial strand, leaving the dermal papilla behind and blocking new hair growth⁵⁶. *Hairless*, which is expressed in the ORS and matrix, is a transcriptional repressor that shares sequence similarities with lysine histone demethylase, an enzyme that removes a positive epigenetic mark from chromatin⁵⁷.

After catagen, follicles lie dormant in a resting phase (telogen). In most mice, the first telogen is short, lasting only 1 or 2 days, from about postnatal day (P)19 to P21 in the mid back. The second telogen, however, lasts more than 2 weeks and begins at about P42 (ref. 37).

Hair follicle stem cells

At the onset of anagen, the cycling portion of the follicle regenerates and undergoes a new round of hair growth. The process necessitates a reservoir of stem cells, which reside in the lowest permanent portion of the follicle. The new downward growth occurs adjacent to the existing club hair, which will eventually be shed. This growth adds a layer to the stem-cell reservoir and causes it to 'bulge', thereby giving this region its name^{58–61} (Fig. 4). Bulge cells divide less frequently than other epithelial cells in the skin, but they are activated at anagen to generate a new hair follicle⁵⁸. The process of bulge-cell activation has been likened to that of embryonic hair follicle morphogenesis. Despite many parallels, there are also differences, and the relationships between the embryonic hair bud and the adult bulge and hair germ are still unfolding.

To understand the special properties underlying bulge stem cells, researchers have turned to microarray profiling, which has led to the identification of about 150 genes that are preferentially expressed in the bulge relative to the proliferating basal cells of the epidermis^{61–63}. Many of these genes — for example, those in the BMP and transforming growth factor- β

pathways — are likely to reflect the special quiescent nature of resting bulge cells, which, in turn, show nuclear localization of phospho-SMAD1 and phospho-SMAD2, respectively^{11,62}. In addition, although bulge cells are Wnt-responsive, as is evident from their expression of TCF3, TCF4 and several Wnt receptor proteins (FZDs), they express proteins typically associated with Wnt inhibition^{21,62}. In this regard, TCF3 can function as a repressor when β -catenin is absent or underrepresented^{19,21}. This could explain how bulge cells maintain an undifferentiated state and why β -catenin stabilization and BMP inhibition are required to activate hair follicle morphogenesis^{12,23–25}. Some of these features have also recently been noted for human bulge stem cells, which could be important for future clinical applications⁶⁴. These parallels are intriguing, because in human skin epidermal stem cells are likely to have the main role in homeostasis — that is, to supply the thick epidermal barrier of the skin surface — whereas in mice, maintaining the protective hair coat is more important.

Several other stimuli function at or slightly downstream of follicle stem-cell activation. One of these is the drug minoxidil. More commonly known as Rogaine, minoxidil was first identified as a vasodilator, and among the cellular responses that have been attributed to its action is prostaglandin synthesis⁵⁴. Cyclosporin A is another hair growth stimulus, which seems to act in promoting anagen or inhibiting catagen^{41,54,56}. Given that the stabilization of β -catenin also promotes anagen, it is interesting to speculate that the pharmacological targets of minoxidil and cyclosporin A might be downstream of β -catenin stabilization. In this regard, it should also be noted that Wnt signalling is only one of several pathways that can lead to β -catenin stabilization, and 'Wnt reporter' activity reflects merely the activation of β -catenin/LEF/TCF-regulated genes.

A number of questions remain. Although grafting and lineage tracing experiments have shown that some bulge cells are stem cells^{61,63,65}, it is still not known whether this is the case for most bulge cells. During the transition to anagen, most bulge cells remain quiescent, and only a small number residing at the base of this segment seem to be activated⁶². Moreover, in the whisker cycle of rodents, the dermal papilla never reaches the bulge before the start of a new anagen. It has been proposed that stem cells from the bulge migrate along the ORS to the base of the follicle, where they become activated⁶⁰. Whether this happens in all hair follicles is not known.

Despite the many questions, it is helpful to place the data at hand in a conceptual model for how stem-cell activation in the hair cycle might be achieved. Upon stem-cell activation, β -catenin/LEF1/TCF becomes activated in the early hair germ and persists at the follicle base throughout anagen. Association with the dermal papilla seems to be critical in accumulating a sufficient level of 'activating factor' to trigger proliferation and/or lineage commitment of these cells. When the dermal papilla moves away from the bulge, the epidermal stem cells return to quiescence. The cells at the base of the follicle that receive sustained dermal papilla stimulation are highly proliferative. Through local differences in their microenvironment they progress to a point of no return, executing terminal lineage programmes to produce the hair and IRS. In this way, the levels of β -catenin/TCF/LEF signalling determine the outcome of the stem cells and their lineages.

Follicle stem-cell maintenance

Once follicle stem cells have been expended, niche vacancies must be replenished. The process seems to be one of self renewal, although stem-cell maintenance is still poorly understood. Balancing *c-myc* expression seems to be crucial, as its overexpression in transgenic mice causes follicle stem cells to proliferate and terminally differentiate, and, conversely, conditional ablation of *c-myc* results in alopecia owing to a long-term failure to sustain follicle stem cells^{66–68}. LHX2 also seems to function in follicle stem-cell maintenance, as exemplified by the fact that LHX2-deficient bulge stem cells are unable to retain label and use stem cells more rapidly than their wild-type counterparts²⁷. Thus, whereas LHX2 seems to function as a molecular brake in regulating the switch between hair follicle stem-cell maintenance and activation, *c-Myc* acts as the accelerator.

Several other factors have been identified as candidates in follicle stem-cell maintenance, including telomerase, cell-division cycle 42

(Cdc42) and Rac1 (reviewed in ref. 36). It has been postulated that Rac1 might help to maintain stem cells by negatively regulating *c-Myc* and by enhancing integrin levels⁶⁹, whereas Cdc42 might control progenitor differentiation through β -catenin stabilization⁷⁰.

Sebaceous gland development

Sebaceous glands are an appendage of the hair follicle, located above the bulge and arrector pili muscle and just below the hair shaft orifice at the skin surface. Sebaceous gland progenitor cells emerge near the conclusion of embryogenesis, but the gland does not mature until just after birth. The main role of the gland is to generate terminally differentiated sebocytes, which produce lipids and sebum. When sebocytes disintegrate, they release these oils into the hair canal for lubrication and protection against bacterial infections. Sebaceous gland homeostasis necessitates a progenitor population of cells that gives rise to a continual flux of proliferating, differentiating and, finally, dead cells that are lost through the hair canal.

The basement membrane that demarcates the mesenchymal–epithelial border also surrounds the gland. Not surprisingly, the sebaceous gland cells that are attached to the inner surface of this membrane share many of the features of epidermal keratinocytes, including proliferative capacity and expression of *c-myc*. Overexpression of *c-myc* induces sebaceous gland hyperplasia at the expense of hair follicle differentiation, which is suggestive of a special role for this transcription factor in the gland^{71,72}. Sebocyte differentiation involves the adipogenic transcription factor peroxisome proliferator-activated receptor- γ (PPAR- γ). PPAR- γ -null mice are not viable, but PPAR- γ -null embryonic stem cells have been shown to contribute poorly to the sebaceous glands of chimaeric mice⁴⁷. In addition, alterations in Wnt, hedgehog and Notch signalling proteins and their effectors result in perturbations in sebaceous gland development, although it is not yet clear how these factors contribute to the sebaceous gland lineage.

Until recently, little was known about how the sebaceous gland is formed and how it is able to sustain its dynamic balance of growth and terminal differentiation. Stem cells have been implicated in the process of sebaceous gland homeostasis, and support for a resident progenitor-cell population comes from fate mapping by retroviral labelling of epithelial cells *in vivo*⁷³. BLIMP1 was recently identified as a marker, and genetic lineage studies revealed the ability of BLIMP1-positive cells to generate sebaceous glands⁴⁷. In culture, BLIMP1-positive cells yield sebocyte colonies and undergo self-renewal. The results of loss-of-function and molecular studies suggest that BLIMP1 functions by repressing *c-myc*, so that as sebocyte progenitors switch off BLIMP1 they become proliferative and initiate sebocyte differentiation⁴⁷.

When *Blimp1* is mutated, or when the skin is injured, the homeostasis of the sebaceous gland is perturbed. Interestingly, under these conditions, bulge cells can be mobilized to maintain the sebaceous gland. Such a precursor–product relationship has been documented by engraftment experiments with isolated bulge stem cells^{61,63,65}, and by analysing the rudimentary follicle structures of either hairless mutant mice⁷⁴ or *Blimp1* conditionally null mice⁴⁷.

Epidermal development

Towards the end of embryonic development, the interfollicular epidermis (IFE) reaches maturity. Epidermis is composed of layers, the outermost of which is at the skin surface. After birth, its role is to guard against infection, to prevent dehydration, and to undergo re-epithelialization after wound injuries. To accomplish these feats, the epidermis constantly replenishes itself by a process of homeostasis. During this process, dividing cells in the innermost (basal) layer continually execute a programme of terminal differentiation, move outwards and are sloughed from the skin surface.

When exposed to physical trauma and chemical assaults the epidermis must also protect itself, which it does by producing copious amounts of cytoplasmic heteropolymers known as intermediate filaments that are composed of keratin proteins. As cells exit from the basal layer and begin their journey towards the skin surface, they switch from the expression of

keratins K14 and K5 to K1 and K10 (ref. 75). This switch was discovered more than 25 years ago, and remains the most reliable indication that an epidermal cell has undergone a commitment to terminally differentiate. The first suprabasal cells are known as spinous cells, reflecting their cytoskeleton of K1/K10 filament bundles connected to robust cell–cell junctions known as desmosomes. These connections provide a cohesive, integrated mechanical framework across and within stacks of epithelial sheets. K6, K16 and K17 are also expressed suprabasally, but only in hyperproliferative situations such as wound healing. This keratin network not only remodels the cytoskeleton for migration but also regulates cell growth through binding to the adaptor protein 14-3-3 σ and stimulating Akt/mTOR (mammalian target of rapamycin) signalling^{76,77}.

As spinous cells progress to the granular layer, they produce electron-dense keratohyalin granules packed with the protein profilaggrin, which, when processed, bundles keratin intermediate filaments even more to generate large macrofibrillar cables. In addition, cornified envelope proteins, which are rich in glutamine and lysine residues, are synthesized and deposited under the plasma membrane of the granular cells. When the cells become permeabilized to calcium, they activate transglutaminase, generating γ -glutamyl ϵ -lysine crosslinks to create an indestructible proteinaceous sac to hold the keratin macrofibrils. The final steps of terminal differentiation involve the destruction of cellular organelles including the nucleus, and the extrusion of lipid bilayers, packaged in lamellar granules, onto the scaffold of the cornified envelope. The dead stratum corneum cells create an impenetrable seal that is continually replenished as inner layer cells move outwards and are sloughed from the skin surface.

Controlling basal cell behaviour

Epidermal growth and proliferation must be carefully balanced: too little proliferation results in thinning of the skin and loss of protection, and too much is a characteristic of hyperproliferative disorders, including psoriasis (see page 866) and cancers. During normal homeostasis the epidermis must be able to sense and replace basal vacancies, and when wounds occur cells must migrate and proliferate but also sense when to stop after wound closure. Finally, the epidermis must have an SOS system to recruit immune cells to fight infection and aid in wound repair, and to turn off the response once the wound has been closed. How does the epidermis achieve all this?

Basal cell proliferation relies on an underlying basement membrane that is rich in extracellular matrix (ECM) proteins and growth factors. Basal cells attach to this structure through two types of cell-junction adhesion complexes, composed of integrins. Hemidesmosomes contain a transmembrane core of $\alpha_6\beta_4$ integrins that connect intracellularly to the keratin intermediate filament network and provide mechanical strength. Focal adhesions contain $\alpha_3\beta_1$ integrins, which connect to the actin and microtubule networks. Both types of junction adhere extracellularly to laminin 5, the main ECM ligand for the epidermal basement membrane. $\alpha\beta_1$ integrins seem to be especially important in assembling and organizing this membrane^{78,79}.

Integrins also function in growth control and migration. Although these growth regulatory circuitries are many and complex, they involve physical interactions with regulatory kinases. To permit migration in response to wounding, cell–substratum junctions must be dynamically turned over. Integrins differ in their relative roles in wound-repair: whereas cells lacking $\alpha\beta_1$ integrins are less migratory, those lacking $\alpha_6\beta_4$ integrins show increased migratory behaviour^{79–81}. The positive role of β_1 integrin in migration seems to result in part from its ability to directly bind and control the activity of focal adhesion kinase (FAK). This is a non-receptor tyrosine kinase that, in turn, functions as a master switch in negatively regulating cellular tension imposed by an elaborate actin–myosin network associated with focal adhesions^{80,81}.

To function as a tissue, basal cells must also adhere to one another. They do this not only by means of desmosomes but also through adherens junctions (Fig. 5). Adherens junctions are composed of a transmembrane core of E-cadherin, which binds two related proteins, β -catenin and p120-catenin⁸². Just as desmosomes and hemidesmosomes form an integral network with intermediate filaments, adherens junctions

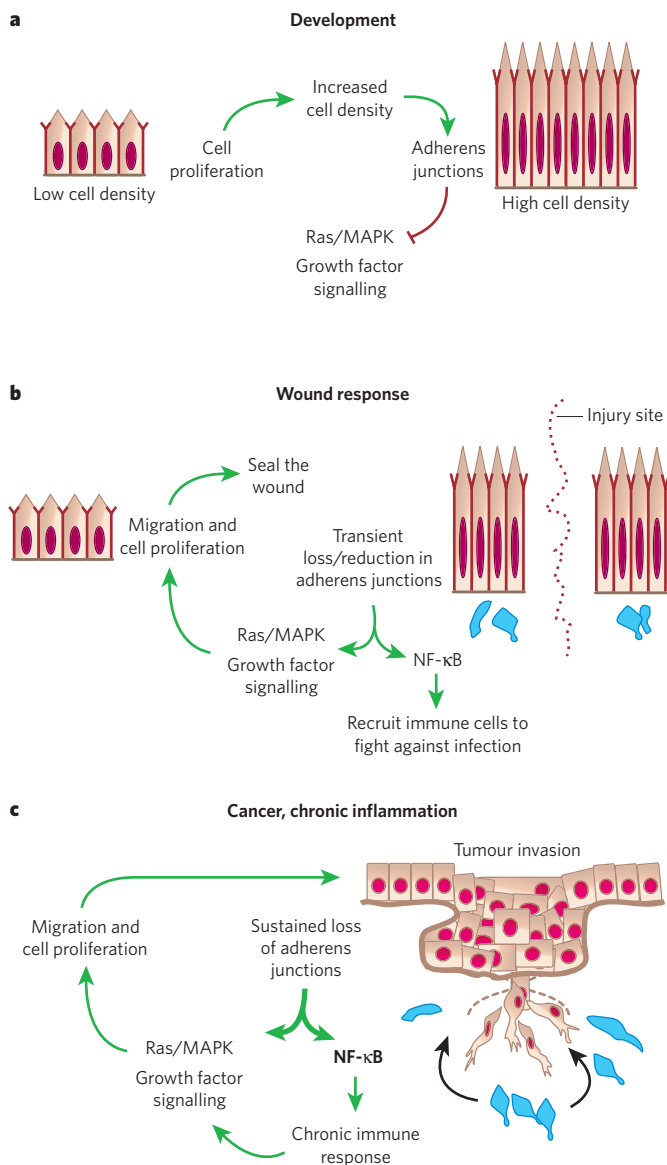


Figure 5 | Model for control of epidermal proliferation. **a**, During development, basal epidermal cell density is low and cells proliferate. As cells increase in density the number of adherens junctions at cell–cell borders increases. This provides a regulatory feedback loop to control Ras–MAPK activity and cell proliferation. **b**, In a wound response, adherens junctions are severed at the wound site, stimulating the Ras–MAPK pathway and triggering a hyperproliferative response and cell migration. Adherens junction protein levels also affect NF- κ B, which, when active, results in the recruitment of immune cells (blue) to protect against infection. After wound repair, regulatory feedback pathways cause these processes to be dampened and the system returns to normal. **c**, Cancer and chronic inflammatory disorders can lead to mutations and/or permanent changes in adherens junction proteins. This results in a break in the regulatory circuitry, and the proliferative and NF- κ B responses become constitutively activated.

and focal adhesions orchestrate actin–myosin dynamics throughout the cells of the tissue. Adherens junctions do so by associating with an array of actin regulatory proteins, which include p120-catenin's interactions with Rho-GTPase proteins and β -catenin's association with α -catenin, an unrelated protein that can bind vasodilator-stimulated phosphoprotein (also known as ENA), formins, ajuba and α -actinin proteins⁸². Exactly how adherens junctions and focal adhesions regulate the actin–myosin network to coordinate adhesion and migration is not yet clear but, typically, whenever cell–cell junctions are reinforced, integrin dynamics are diminished. This inverse regulation is likely to be essential during

wound repair, in which epidermal outgrowth must occur in a polarized and orchestrated manner.

Recently, researchers have begun to realize that cellular junctions act as signalling centres and have functions that extend beyond their classical roles in cell adhesion and cytoskeletal dynamics. The longest standing example is β -catenin, which is now well known for its dual role in adhesion and transcription. p120-Catenin also functions not only in regulating Rho GTPases and adherens junction assembly, but also as a transcriptional co-factor⁸³. In addition, through a mechanism that is not fully understood but seems to involve the small Rho GTPases, p120-catenin and α -catenin can affect the transcriptional status of NF- κ B, which, in turn, governs the production and secretion of cytokines and chemokines and the recruitment of immune cells^{84,85}. The abrogation of JunB, a downstream inhibitor of EGF signalling, also triggers chemokine/cytokine expression⁸⁶.

When sustained activation of NF- κ B arises — for example, through dysfunction of p120-catenin, α -catenin and/or JunB — signalling from the epidermis to immune cells goes awry. This triggers a molecular war between the epidermis, which spews out cytokines and chemokines, and the recruited immune cells, which respond in a similar way and send proliferative signals to the epidermis. Further studies will be needed to ascertain the controls on these normal pathways for the repair of the epidermis in response to injury, stress and infection, and to elucidate the extent to which these various pathways intersect in recruiting immune cells to the skin. NF- κ B regulation is particularly important because it also mediates a TNFR1-dependent hypoproliferative influence on healthy epidermis⁸⁷.

The consequences of α -catenin loss are particularly severe, and the epidermis progresses to a condition resembling invasive squamous-cell carcinoma²⁹. Although the mechanisms are not yet fully understood, they are likely to involve an upregulation in integrins and tyrosine kinase activity^{85,88,89}. An emerging view of adherens junctions is that they serve as molecular sensors for 'crowd control' and wound repair. In this model, when basal cell density is low or when the epidermis is severed, as in a wound, a reduction in adherens junctions triggers cell migration and proliferation, and immune cells are recruited to protect against possible infection. Once proper cell density has been achieved and adherens junction formation becomes optimal, the system returns to normal (Fig. 5). When this circuitry is defective — for example, when α -catenin or p120 is mutated — the system becomes imbalanced and chronic inflammation and hyperproliferative disorders, including cancer, can result.

Maintaining and controlling the epidermis

Interfollicular skin epithelium is organized into columns of hexagonally packed cells. This organization has been revealed by genetic lineage tracing involving either retrovirally infected, β -galactosidase-expressing cultured keratinocytes grafted onto immunodeficient mice or direct infection of dermabraded skin with a β -galactosidase-expressing virus^{73,90,91}. Initial discrete stacks of β -galactose-positive cells expand to about 10 adjacent stacks over time, which suggests that epidermal proliferative units (EPU)s may be composed of one basal stem cell that occasionally divides laterally to produce cells with a more transient proliferative capacity^{92,93}. Irrespective of the relative proliferative capacity of individual basal cells, their ability to fuel a constant flux of terminally differentiating cells places the epidermis in a constant state of dynamic equilibrium. In humans, the epidermis replenishes itself every 4 weeks throughout life.

The ability of epidermal domains to be maintained individually can be explained by one of two different mechanisms, which are not necessarily mutually exclusive (Fig. 6). *In vitro* studies support a delamination model, whereby epidermal cells with the highest integrin levels also have the highest proliferative and attachment potential, whereas those with low integrin levels detach and terminally differentiate⁹². Studies on embryonic skin development in mice have revealed that concomitantly with stratification, basal cells can also orient their spindle poles asymmetrically relative to the basement membrane^{40,94}. Such divisions seem to be asymmetric, leading to one proliferative basal daughter and one detached, suprabasal daughter. Postnatally, cellular divisions in black skin are markedly reduced, but in other areas or other adult tissues in which the epithelium is thicker, asymmetrically oriented spindles are apparent.

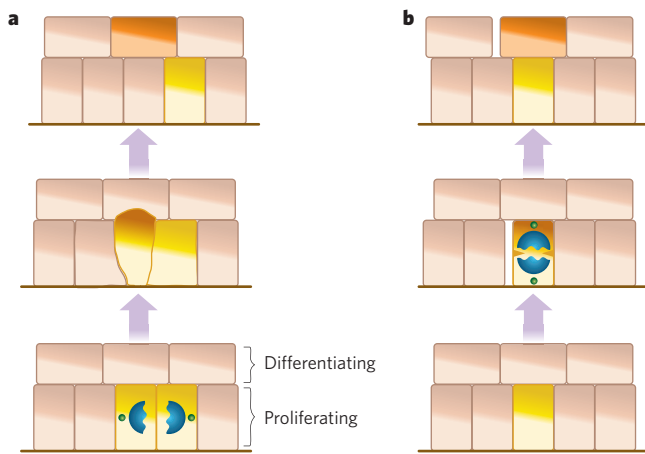


Figure 6 | Models of how epidermal homeostasis might be achieved.

Symmetric and asymmetric cell division have been observed in the epidermis, and both are likely to contribute to epidermal homeostasis. Models of how they might do so are shown. Both are predicated on the basis of cellular attachment to an underlying basement membrane providing the main means by which basal epidermal progenitor cells are able to preserve their mitotic capability. **a**, Symmetric division is thought to supply new cells through stratification (the formation of layers) and differentiation by delamination (the detachment of a cell from the basement membrane). **b**, Asymmetric division is believed to contribute through stratification and differentiation by asymmetric partitioning of cellular contents into the two daughter cells.

Interestingly, mitotic basal epidermal cells with perpendicularly aligned spindles seem to use some of the same proteins involved in an ancient mechanism for asymmetrically partitioning the cellular contents and fates of two daughter cells in *Drosophila* neuroblasts⁴⁰. It is tempting to speculate that the attached basal daughter might preferentially receive integrins, growth factor receptors and other key growth promoting machinery at the base of the cell. An additional factor involved is the basal transcription factor p63, which not only promotes proliferation in the epidermis but also seems to be required for epidermal stratification^{95,96}. In the absence of p63, basal cells seem to divide only symmetrically⁴⁰.

In the fly neuroblast, the asymmetric division machinery preferentially partitions Notch signalling. Although it is not yet clear whether this feature is used by the epidermis, it is intriguing that suprabasal epidermal cells use NICD, the transcriptional effector of Notch signalling, to regulate their differentiation^{51,52}. Additional transcriptional regulators of epidermal differentiation include p21, AP2, C/EBP (CCAAT/enhancer binding protein), Kruppel-like factor and PPAR proteins⁴⁴. Another important factor in epidermal stratification and differentiation is IKK- α (inhibitor of NF- κ B kinase- α), which exerts its action on the epidermis in an NF- κ B-independent manner⁹⁷. By a mechanism not yet determined, the polarity established by the basement membrane, cell substratum and cell–cell junctions generates the epidermal architecture, which, in turn, governs the transcriptional balance between proliferation and differentiation.

The future of skin stem cells in regenerative medicine.

Through understanding the basics of its biology, we now know the skin to be a rich source of readily accessible stem cells. Although the main purpose of the stem cells in the bulge is to maintain homeostasis of the hair follicle during cycling, bulge cells can be mobilized to replenish the cells of the sebaceous gland and the epidermis when necessary^{17,47,59,61,63,65,91}. Whether resident progenitor cells in the basal layer of the epidermis, the sebaceous gland and the follicle matrix can also select alternative lineages when required remains unknown. Although in normal skin the molecular steps in these lineages are becoming increasingly well defined both temporally and spatially, their potential for re-routing their lineage when exposed to a new microenvironment, for example in wounding, engraftment or human disease, is less clear.

The level of plasticity afforded these stem cells is becoming increasingly

important as the potential of stem cells in regenerative medicine continues to be explored. Developmental biologists have long been aware that embryonic epithelial cells are sensitive to the mesenchyme to which they are exposed. When confronted with mesenchyme from chick wing, epidermis from the leg produces feathers, and, similarly, mesenchyme from leg can prompt wing epidermis to make scales. These special features seem to be lost in most cells as development proceeds, but do multipotent stem cells of the skin retain this potential? If bulge stem cells exposed to corneal mesenchyme produce hair follicles, this would not be helpful, but if they produce cornea, they could be used to treat certain types of blindness. Other possible uses for bulge stem cells might be in treating chronic ulcers or hair disorders. The potential of cultured basal epidermal keratinocytes has already been realized for treating burns patients⁹⁸ (see page 874).

On the basis of studies on haematopoietic stem cells, the promise of bulge cells seems slim for generating non-epithelial tissues. That said, the results of recent studies suggest that bulge-cell nuclei from adult mice can be reprogrammed when placed in an enucleated, unfertilized oocyte, and cloned mice have been generated with such technology⁹⁹. If current ethical and technical hurdles can be overcome to permit cloning of human hybrid embryonic stem cells from oocytes and adult skin nuclei, then it might be possible to harness their potential for broader uses in regenerative medicine in the future. A final note in closing is that the accessibility of the skin as a potential source of stem cells has not gone unnoticed in the broader research community. Given that there are many different cell types in the skin, the recent identification of melanoblast stem cells, mesenchymal stem cells and neural-like stem cells seem likely to be the tip of a promising iceberg for future explorations. ■

1. Stern, C. D. Neural induction: old problem, new findings, yet more questions. *Development* **132**, 2007–2021 (2005).
2. M'Boneko, V. & Merker, H. J. Development and morphology of the periderm of mouse embryos (days 9–12 of gestation). *Acta Anat. (Basel)* **133**, 325–336 (1988).
3. Atit, R. et al. β -catenin activation is necessary and sufficient to specify the dorsal dermal fate in the mouse. *Dev. Biol.* **296**, 164–176 (2006).
4. Hardy, M. H. The secret life of the hair follicle. *Trends Genet.* **8**, 55–61 (1992).
5. Olivera-Martinez, I., Thelu, J. & Dhauailly, D. Molecular mechanisms controlling dorsal dermis generation from the somitic dermatome. *Int. J. Dev. Biol.* **48**, 93–101 (2004).
6. Davidson, D. The mechanism of feather pattern development in the chick. II. Control of the sequence of pattern formation. *J. Embryol. Exp. Morphol.* **74**, 261–273 (1983).
7. Petitot, A. et al. A crucial role for Fgf2-Ilb signalling in epidermal development and hair follicle patterning. *Development* **130**, 5493–5501 (2003).
8. Jung, H. S. et al. Local inhibitory action of BMPs and their relationships with activators in feather formation: implications for periodic patterning. *Dev. Biol.* **196**, 11–23 (1998).
9. Noramly, S. & Morgan, B. A. BMPs mediate lateral inhibition at successive stages in feather tract development. *Development* **125**, 3775–3787 (1998).
10. Botchkarev, V. A. et al. Noggin is a mesenchymally derived stimulator of hair-follicle induction. *Nature Cell Biol.* **1**, 158–164 (1999).
11. Mou, C., Jackson, B., Schneider, P., Overbeek, P. A. & Headon, D. J. Generation of the primary hair follicle pattern. *Proc. Natl Acad. Sci. USA* **103**, 9075–9080 (2006).
12. Gat, U., DasGupta, R., Degenstein, L. & Fuchs, E. *De novo* hair follicle morphogenesis and hair tumors in mice expressing a truncated β -catenin in skin. *Cell* **95**, 605–614 (1998).
13. DasGupta, R. & Fuchs, E. Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. *Development* **126**, 4557–4568 (1999).
14. Noramly, S., Freeman, A. & Morgan, B. A. β -catenin signaling can initiate feather bud development. *Development* **126**, 3509–3521 (1999).
15. St-Jacques, B. et al. Sonic hedgehog signaling is essential for hair development. *Curr. Biol.* **8**, 1058–1068 (1998).
16. Oro, A. E. & Higgins, K. Hair cycle regulation of Hedgehog signal reception. *Dev. Biol.* **255**, 238–248 (2003).
17. Levy, V., Lindon, C., Harfe, B. D. & Morgan, B. A. Distinct stem cell populations regenerate the follicle and interfollicular epidermis. *Dev. Cell* **9**, 855–861 (2005).
18. Reddy, S. et al. Characterization of *Wnt* gene expression in developing and postnatal hair follicles and identification of *Wnt5a* as a target of Sonic hedgehog in hair follicle morphogenesis. *Mech. Dev.* **107**, 69–82 (2001).
19. Merrill, B. J., Gat, U., DasGupta, R. & Fuchs, E. Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. *Genes Dev.* **15**, 1688–1705 (2001).
20. Hurlstone, A. & Clevers, H. T-cell factors: turn-ons and turn-offs. *EMBO J.* **21**, 2303–2311 (2002).
21. Nguyen, H., Rendl, M. & Fuchs, E. Tcf3 maintains stem cells and represses cell fate determination in skin. *Cell* **127**, 171–183 (2006).
22. van Genderen, C. et al. Development of several organs that require inductive epithelial–mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev.* **8**, 2691–2703 (1994).
23. Van Mater, D., Kolligs, F. T., Dlugosz, A. A. & Fearon, E. R. Transient activation of β -catenin signaling in cutaneous keratinocytes is sufficient to trigger the active growth phase of the hair cycle in mice. *Genes Dev.* **17**, 1219–1224 (2003).
24. Lo Celso, C., Prowse, D. M. & Watt, F. M. Transient activation of β -catenin signalling in adult mouse epidermis is sufficient to induce new hair follicles but continuous activation is required to maintain hair follicle tumours. *Development* **131**, 1787–1799 (2004).

25. Lowry, W. E. *et al.* Defining the impact of β -catenin/Tcf transactivation on epithelial stem cells. *Genes Dev.* **19**, 1596–1611 (2005).
26. Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G. & Birchmeier, W. β -Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* **105**, 533–545 (2001).
27. Andl, T., Reddy, S. T., Gaddapara, T. & Millar, S. E. WNT signals are required for the initiation of hair follicle development. *Dev. Cell* **2**, 643–653 (2002).
28. Jamora, C., DasGupta, R., Kocieniewski, P. & Fuchs, E. Links between signal transduction, transcription and adhesion in epithelial bud development. *Nature* **422**, 317–322 (2003).
29. Kobiela, K., Pasolli, H. A., Alonso, L., Polak, L. & Fuchs, E. Defining BMP functions in the hair follicle by conditional ablation of BMP receptor 1A. *J. Cell Biol.* **163**, 609–623 (2003).
30. Andl, T. *et al.* Epithelial Bmp1a regulates differentiation and proliferation in postnatal hair follicles and is essential for tooth development. *Development* **131**, 2257–2268 (2004).
31. Rhee, H., Polak, L. & Fuchs, E. Lhx2 maintains stem cells character in hair follicles. *Science* **312**, 1946–1949 (2006).
32. Headon, D. J. *et al.* Gene defect in ectodermal dysplasia implicates a death domain adapter in development. *Nature* **414**, 913–916 (2001).
33. Schmidt-Ullrich, R. *et al.* NF- κ B transmits Eda A1/EdaR signalling to activate Shh and cyclin D1 expression, and controls post-initiation hair placode down growth. *Development* **133**, 1045–1057 (2006).
34. Rinn, J. L., Bondre, C., Gladstone, H. B., Brown, P. O. & Chang, H. Y. Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet.* **2**, e119 (2006).
35. Atit, R., Conlon, R. A. & Niswander, L. EGF signaling patterns the feather array by promoting the interbud fate. *Dev. Cell* **4**, 231–240 (2003).
36. Blanpain, C. & Fuchs, E. Epidermal stem cells of the skin. *Annu. Rev. Cell Dev. Biol.* **22**, 339–373 (2006).
37. Muller-Rover, S. *et al.* A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. *J. Invest. Dermatol.* **117**, 3–15 (2001).
38. Kopan, R. *et al.* Genetic mosaic analysis indicates that the bulb region of coat hair follicles contains a resident population of several active multipotent epithelial lineage progenitors. *Dev. Biol.* **242**, 44–57 (2002).
39. Legue, E. & Nicolas, J. F. Hair follicle renewal: organization of stem cells in the matrix and the role of stereotyped lineages and behaviors. *Development* **132**, 4143–4154 (2005).
40. Lechler, T. & Fuchs, E. Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* **437**, 275–280 (2005).
41. Langbein, L. & Schweizer, J. Keratins of the human hair follicle. *Int. Rev. Cytol.* **243**, 1–78 (2005).
42. Chan, E. F., Gat, U., McNiff, J. M. & Fuchs, E. A common human skin tumour is caused by activating mutations in β -catenin. *Nature Genet.* **21**, 410–413 (1999).
43. Godwin, A. R. & Capocchi, M. R. Hoxc13 mutant mice lack external hair. *Genes Dev.* **12**, 11–20 (1998).
44. Dai, X. & Segre, J. A. Transcriptional control of epidermal specification and differentiation. *Curr. Opin. Genet. Dev.* **14**, 485–491 (2004).
45. Ellis, T. *et al.* The transcriptional repressor CDP (Cut1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev.* **15**, 2307–2319 (2001).
46. Kaufman, C. K. *et al.* GATA-3: an unexpected regulator of cell lineage determination in skin. *Genes Dev.* **17**, 2108–2122 (2003).
47. Horsley, V. *et al.* Blimp1 defines a progenitor population that governs cellular input to the sebaceous gland. *Cell* **126**, 597–609 (2006).
48. Ancelin, K. *et al.* Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nature Cell Biol.* **8**, 623–630 (2006).
49. Chang, D. H., Angelin-Duclos, C. & Calame, K. BLIMP-1: trigger for differentiation of myeloid lineage. *Nature Immunol.* **1**, 169–176 (2000).
50. De Strooper, B. *et al.* A presenilin-1-dependent γ -secretase-like protease mediates release of Notch intracellular domain. *Nature* **398**, 518–522 (1999).
51. Pan, Y. *et al.* γ -Secretase functions through Notch signaling to maintain skin appendages but is not required for their patterning or initial morphogenesis. *Dev. Cell* **7**, 731–743 (2004).
52. Blanpain, C., Lowry, W. E., Pasolli, H. A. & Fuchs, E. Canonical Notch signaling functions as a commitment switch in the epidermal lineage. *Genes Dev.* **20**, 3022–3035 (2006).
53. Ming Kwan, K., Li, A. G., Wang, X. J., Wurst, W. & Behringer, R. R. Essential roles of BMP1-IA signaling in differentiation and growth of hair follicles and in skin tumorigenesis. *Genesis* **39**, 10–25 (2004).
54. Schmidt-Ullrich, R. & Paus, R. Molecular principles of hair follicle induction and morphogenesis. *BioEssays* **27**, 247–261 (2005).
55. Tong, X. & Coulombe, P. A. Keratin 17 modulates hair follicle cycling in a TNF α -dependent fashion. *Genes Dev.* **20**, 1353–1364 (2006).
56. Potter, G. B. *et al.* The hairless gene mutated in congenital hair loss disorders encodes a novel nuclear receptor corepressor. *Genes Dev.* **15**, 2687–2701 (2001).
57. Klose, R. J. *et al.* The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442**, 312–316 (2006).
58. Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61**, 1329–1337 (1990).
59. Taylor, G., Lehrer, M. S., Jensen, P. J., Sun, T. T. & Lavker, R. M. Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* **102**, 451–461 (2000).
60. Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K. & Barrandon, Y. Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* **104**, 233–245 (2001).
61. Blanpain, C., Lowry, W. E., Geoghegan, A., Polak, L. & Fuchs, E. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* **118**, 635–648 (2004).
62. Tumber, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
63. Morris, R. J. *et al.* Capturing and profiling adult hair follicle stem cells. *Nature Biotechnol.* **22**, 411–417 (2004).
64. Ohshima, M. *et al.* Characterization and isolation of stem cell-enriched human hair follicle bulge cells. *J. Clin. Invest.* **116**, 249–260 (2006).
65. Claudinot, S., Nicolas, M., Oshima, H., Rochat, A. & Barrandon, Y. Long-term renewal of hair follicles from clonogenic multipotent stem cells. *Proc. Natl Acad. Sci. USA* **102**, 14677–14682 (2005).
66. Arnold, I. & Watt, F. M. c-Myc activation in transgenic mouse epidermis results in mobilization of stem cells and differentiation of their progeny. *Curr. Biol.* **11**, 558–568 (2001).
67. Waikel, R. L., Kawachi, Y., Waikel, P. A., Wang, X. J. & Roop, D. R. Deregulated expression of c-Myc depletes epidermal stem cells. *Nature Genet.* **28**, 165–168 (2001).
68. Zanet, J. *et al.* Endogenous Myc controls mammalian epidermal cell size, hyperproliferation, endoreplication and stem cell amplification. *J. Cell Sci.* **118**, 1693–1704 (2005).
69. Benitah, S. A., Frye, M., Glogauer, M. & Watt, F. M. Stem cell depletion through epidermal deletion of Rac1. *Science* **309**, 933–935 (2005).
70. Wu, X. *et al.* Cdc42 controls progenitor cell differentiation and β -catenin turnover in skin. *Genes Dev.* **20**, 571–585 (2006).
71. Niemann, C., Owens, D. M., Hulsken, J., Birchmeier, W. & Watt, F. M. Expression of Δ Nle1 in mouse epidermis results in differentiation of hair follicles into squamous epidermal cysts and formation of skin tumours. *Development* **129**, 95–109 (2002).
72. Takeda, H. *et al.* Human sebaceous tumors harbor inactivating mutations in LEF1. *Nature Med.* **12**, 395–397 (2006).
73. Ghazizadeh, S. & Taichman, L. B. Multiple classes of stem cells in cutaneous epithelium: a lineage analysis of adult mouse skin. *EMBO J.* **20**, 1215–1222 (2001).
74. Pantelejev, A. A., Paus, R. & Christiano, A. M. Patterns of hairless (*hr*) gene expression in mouse hair follicle morphogenesis and cycling. *Am. J. Pathol.* **157**, 1071–1079 (2000).
75. Fuchs, E. & Green, H. Changes in keratin gene expression during terminal differentiation of the keratinocyte. *Cell* **19**, 1033–1042 (1980).
76. Coulombe, P. A. & Wong, P. Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds. *Nature Cell Biol.* **6**, 699–706 (2004).
77. Kim, S., Wong, P. & Coulombe, P. A. A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. *Nature* **441**, 362–365 (2006).
78. Watt, F. M. Role of integrins in regulating epidermal adhesion, growth and differentiation. *EMBO J.* **21**, 3919–3926 (2002).
79. Wilhelmson, K., Litjens, S. H. & Sonnenberg, A. Multiple functions of the integrin α 6 β 4 in epidermal homeostasis and tumorigenesis. *Mol. Cell Biol.* **26**, 2877–2886 (2006).
80. Raghavan, S., Vaezi, A. & Fuchs, E. A role for α 6 β 1 integrins in focal adhesion function and polarized cytoskeletal dynamics. *Dev. Cell* **5**, 415–427 (2003).
81. Brakebusch, C. & Fassler, R. β 1 integrin function in vivo: adhesion, migration and more. *Cancer Metastasis Rev.* **24**, 403–411 (2005).
82. Perez-Moreno, M. & Fuchs, E. Catenins: keeping cells from getting their signals crossed. *Dev. Cell* **11**, 601–612 (2006).
83. van Roy, F. M. & McCrea, P. D. A role for Kaiso-p120ctn complexes in cancer? *Nature Rev. Cancer* **5**, 956–964 (2005).
84. Perez-Moreno, M. *et al.* p120-catenin mediates inflammatory responses in the skin. *Cell* **124**, 631–644 (2006).
85. Kobiela, K. & Fuchs, E. Links between α -catenin, NF- κ B, and squamous cell carcinoma in skin. *Proc. Natl Acad. Sci. USA* **103**, 2322–2327 (2006).
86. Zenz, R. *et al.* Psoriasis-like skin disease and arthritis caused by inducible epidermal deletion of Jun proteins. *Nature* **437**, 369–375 (2005).
87. Zhang, J. Y., Green, C. L., Tao, S. & Khavari, P. A. NF- κ B RelA opposes epidermal proliferation driven by TNFR1 and JNK. *Genes Dev.* **18**, 17–22 (2004).
88. Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B. & Fuchs, E. Hyperproliferation and defects in epithelial polarity upon conditional ablation of α -catenin in skin. *Cell* **104**, 605–617 (2001).
89. Zhang, W. *et al.* E-cadherin loss promotes the initiation of squamous cell carcinoma invasion through modulation of integrin-mediated adhesion. *J. Cell Sci.* **119**, 283–291 (2006).
90. Kolodka, T. M., Garlick, J. A. & Taichman, L. B. Evidence for keratinocyte stem cells in vitro: long term engraftment and persistence of transgene expression from retrovirus-transduced keratinocytes. *Proc. Natl Acad. Sci. USA* **95**, 4356–4361 (1998).
91. Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Med.* **11**, 1351–1354 (2005).
92. Mackenzie, I. C. Relationship between mitosis and the ordered structure of the stratum corneum in mouse epidermis. *Nature* **226**, 653–655 (1970).
93. Potten, C. S. Cell replacement in epidermis (keratopoiesis) via discrete units of proliferation. *Int. Rev. Cytol.* **69**, 271–318 (1981).
94. Smart, I. H. Variation in the plane of cell cleavage during the process of stratification in the mouse epidermis. *Br. J. Dermatol.* **82**, 276–282 (1970).
95. Mills, A. A. *et al.* p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* **398**, 708–713 (1999).
96. Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).
97. Hu, Y. *et al.* IKK α controls formation of the epidermis independently of NF- κ B. *Nature* **410**, 710–714 (2001).
98. Green, H. Cultured cells for the treatment of disease. *Sci. Am.* **265**, 96–102 (1991).
99. Li, J., Greco, V., Guasch, G., Fuchs, E. & Mombaerts, P. Mice cloned from adult skin cells. *Proc. Natl Acad. Sci. USA* (in the press).
100. Paus, R. & Cotsarelis, G. The biology of hair follicles. *N. Engl. J. Med.* **341**, 491–497 (1999).

Acknowledgments I am grateful to my many colleagues who helped to establish the groundwork for this review. I also thank J.-F. Nicolas (Pasteur Institute), H. A. Pasolli, H. Rhee and T. Lechler for their helpful suggestions about figures for this review. I am especially indebted to my former mentor, H. Green, his former postdoctoral researchers — T.-T. Sun, J. Rheinwald, F. Watt and Y. Barrandon — and to the members of my laboratory, past and present, all of whom have contributed so heavily to the field of skin biology and who have served as an enormous source of inspiration to my own contributions. E.F. is an Investigator of the Howard Hughes Medical Institute. This work was supported in part by the National Institutes of Health and the Stem Cell Research Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (fuchslb@rockefeller.edu).

Melanocyte biology and skin pigmentation

Jennifer Y. Lin^{1,2} & David E. Fisher²

Melanocytes are phenotypically prominent but histologically inconspicuous skin cells. They are responsible for the pigmentation of skin and hair, and thereby contribute to the appearance of skin and provide protection from damage by ultraviolet radiation. Pigmentation mutants in various species are highly informative about basic genetic and developmental pathways, and provide important clues to the processes of photoprotection, cancer predisposition and even human evolution. Skin is the most common site of cancer in humans. Continued understanding of melanocyte contributions to skin biology will hopefully provide new opportunities for the prevention and treatment of skin diseases.

Melanocytes can absorb ultraviolet radiation (UVR) and survive considerable genotoxic stress. The skin is the main barrier to the external environment, and relies on melanocytes to provide, among other things, photoprotection and thermoregulation by producing melanin. The degree of pigment production manifests as skin 'phototype' (skin colour and ease of tanning)¹ and is the most useful predictor of human skin cancer risk in the general population.

The colours we see in feathers, fur and skin are largely determined by melanocytes. In addition to carotenoids and haemoglobin, melanin is the main contributor to pigmentation. There are two main types of melanin — red/yellow pheomelanin and brown/black eumelanin. Melanin-containing granules are known as melanosomes and are exported from melanocytes to adjacent keratinocytes, where most pigment is found. As a result, pigmentation differences can arise from variation in the number, size, composition and distribution of melanosomes, whereas melanocyte numbers typically remain relatively constant (Fig. 1a, b).

Mutations affecting pigmentation have been identified in many species because they are easily recognizable. Such mutants can be categorized into four groups: hypopigmentation and hyperpigmentation, with or without altered melanocyte number. These phenotypic distinctions have afforded the opportunity to easily classify genes affecting the melanocyte lineage, with respect to viability or differentiation (or both). Some of these mutants function in non-cell-autonomous manner, thereby further revealing cell-cell communication pathways of physiological importance. Collectively, pigmentation or coat colour mutants have become an invaluable resource for the analysis of melanocyte differentiation and as a model for the broader fields of neural-crest development and mammalian genetics.

There are two discrete melanocytic populations in hair follicles: melanocyte stem cells and their differentiated progeny, which reside in geographically distinct locations to comprise a follicular unit that is tightly linked to the surrounding keratinocyte population. Hair follicle melanocyte stem cells have important roles in both normal hair pigmentation and senile hair greying, and specific genetic defects have shed further light on the survival properties of this cell population.

This review summarizes how pigmentation is regulated at the molecular level and how the tanning response provides protection against damage and skin cancer. We discuss recent advances in our knowledge of the genes involved in these processes and how they affect skin and hair colour. We also cover the developmental origin of melanocytes and how they are maintained by melanoblast stem cells, whose eventual depletion may contribute to hair greying. Finally, we detail some questions that research into melanocyte biology hopes to address in the future.

Regulation of pigmentation

Melanocortin-1 receptor

The contribution of melanocytes to pigmentation is conserved through many species. In certain species such as fish, pigment is provided by other cell types, known as xanthophores and iridophores² (Fig. 1c). Despite the identification of more than 100 loci involved in vertebrate pigmentation, the melanocortin-1 receptor (MC1R) is consistently a representative locus and major determinant of pigment phenotype³. The extension locus (*Mc1r^{ext}*) was first identified in mice on the basis of altered coat colour^{4,5}. The recessive mutants have yellow or pheomelanotic hair, whereas wild-type mice have black/brown eumelanotic hair. This mutation has been conserved in furred animals from mammoths to present-day cats and dogs⁶ (Fig. 1d). Other melanocortin family member receptors are found on various cell lineages — for example, *MC4R* is expressed in the hypothalamus, where it modulates energy metabolism⁷.

MC1R encodes a seven-transmembrane, G-protein-coupled receptor. Agonist-bound MC1R activates adenylyl cyclase, inducing cyclic AMP production⁸, which leads to phosphorylation of cAMP responsive-element-binding protein (CREB) transcription factor family members. CREB, in turn, transcriptionally activates various genes, including that encoding microphthalmia transcription factor (*MITF*), the transcription factor that is pivotal to the expression of numerous pigment enzymes and differentiation factors⁹ (Fig. 2). Agonists of human MC1R include α -melanocyte-stimulating hormone (α -MSH) and adrenocorticotrophic hormone (ACTH), and these cause an increase in eumelanin production through elevated cAMP levels^{10,11}. The agouti (*Asip*) gene encodes an antagonist of MC1R¹², which is responsible for the pheomelanotic banding pattern of wild-type mouse fur. An inactivating mutation (non-agouti) at this locus is responsible for the black fur of the C57BL6 mouse strain. Recently, evidence has been reported for an association between a single nucleotide polymorphism in the 3' untranslated region of the human agouti protein and dark hair and brown eyes¹³.

The role of MC1R in hair pigmentation is striking. The human MC1R coding region is highly polymorphic with at least 30 allelic variants, most of which result in a single amino-acid substitution¹⁴. Certain substitutions, such as R151C, R160W and D294H, are associated with red hair. The 'red-head' phenotype is defined not only by hair colour but also by fair skin, inability to tan and a propensity to freckle. Functional studies suggest that these variants encode hypomorphic mutants that are unable to either bind ligand or activate adenylyl cyclase^{15,16}. Thus, it may be possible to have an additive effect among two variant alleles¹⁷. Two point mutations in the second transmembrane domain have yielded constitutively active

¹Harvard Combined Program in Dermatology, Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts 02115, USA. ²Melanoma Program and Department of Pediatric Hematology and Oncology, Dana-Farber Cancer Institute, Children's Hospital Boston, 44 Binney Street, Boston, Massachusetts 02115, USA.

receptors with a resulting dominant dark coat in mice⁵, although such gain-of-function mutations have not yet been reported in humans¹⁸.

Tracing *MC1R* loci through different skin types and geographic regions has led to different theories on the evolution of human pigmentation. Epidemiological studies suggest that pigmentation is under functional constraint in Africa and that this constraint has been lost in the populations that left the African continent. It is not clear whether the drive for selection was necessitated by UVR-induced vitamin D production over protection from the DNA damage caused by UVR or as a result of an undiscovered critical pathway^{14,18}. *MC1R* might be evolutionarily significant for other biological reasons, such as increased κ -opioid analgesia, which was recently linked to variant *MC1R* alleles in both mice and humans¹⁹. Despite the apparently strong influence of *MC1R* on both hair and skin pigmentation, it is clear that other factors are also involved in the control of skin pigmentation, because there are many fair-skinned but dark-haired individuals in whom *MC1R* alone is unlikely to limit skin pigmentation.

Recently *SLC24A5*, which encodes a putative cation exchanger, was identified as the human homologue of a zebrafish gene that causes the 'golden' phenotype²⁰ (Fig. 1). Although its function is unclear, it is plausible that cation chemistry might modulate melanosomal maturation processes. Humans have two primary *SLC24A5* alleles, which differ by a single amino-acid substitution. In almost all Africans and Asians the substitution is alanine, but in 98% of Europeans the allele encodes threonine. The function of *SLC24A5* in human pigmentation remains to be determined, but the correlation of its variants to human populations is striking and suggests that it is important in the control of cutaneous pigmentation.

Melanosomes and melanogenesis

Melanin production occurs predominantly in a lysosome-like structure known as the melanosome. Pheomelanin and eumelanin differ not only in colour but also in the size, shape and packaging of their granules²¹.

Both melanins derive from a common tyrosinase-dependent pathway with the same precursor, tyrosine. The obligatory step is hydroxylation of tyrosine to dopaquinone, from which L-DOPA can also be derived²². The absence or severe dysfunction of tyrosinase and other key pigment enzymes (including *P* gene, the human homologue of the mouse pink-eyed dilution locus, tyrosinase-related protein 1, TRP1, and membrane-associated transporter protein, MATP) results in oculocutaneous albinism (OCA1–4), which presents with intact melanocytes but inability to make pigment (see ref. 23 for a review).

From dopaquinone, the eumelanin and pheomelanin pathways diverge. Two enzymes crucial to eumelanogenesis are the tyrosinase-related proteins TRP1 (also known as GP75 or *b*-locus) and TRP2 (also known as dopachrome tautomerase, DCT). TRP1 and 2 share 40–45% identity with tyrosinase and are useful markers of differentiation. Pheomelanin is derived from conjugation by thiol-containing cysteine or glutathione. As a result, pheomelanin is more photolabile and can produce, among its by-products, hydrogen peroxide, superoxide and hydroxyl radicals, all known triggers of oxidative stress, which can cause further DNA damage. Individual melanocytes typically synthesize both eumelanins and pheomelanins, with the ratio of the two being determined by a balance of variables, including pigment enzyme expression and the availability of tyrosine and sulphhydryl-containing reducing agents in the cell²².

Melanin is packaged and delivered to keratinocytes by melanosomes. The formation, maturation and trafficking of melanosomes is crucial to pigmentation, and defects in this process lead to depigmented and dilutionary disorders such as Hermansky–Pudlak Syndrome (HPS) and Chediak–Higashi Syndrome (CHS)²⁴. On the keratinocyte side, the protease-activated receptor-2 (PAR2), a seven-transmembrane receptor on keratinocytes, has a central role in melanosome transfer²⁵. Once in keratinocytes, melanosomes are distributed and, in response to UVR, positioned strategically over the 'sun-exposed' side of nuclei to form cap-like structures resembling umbrellas.

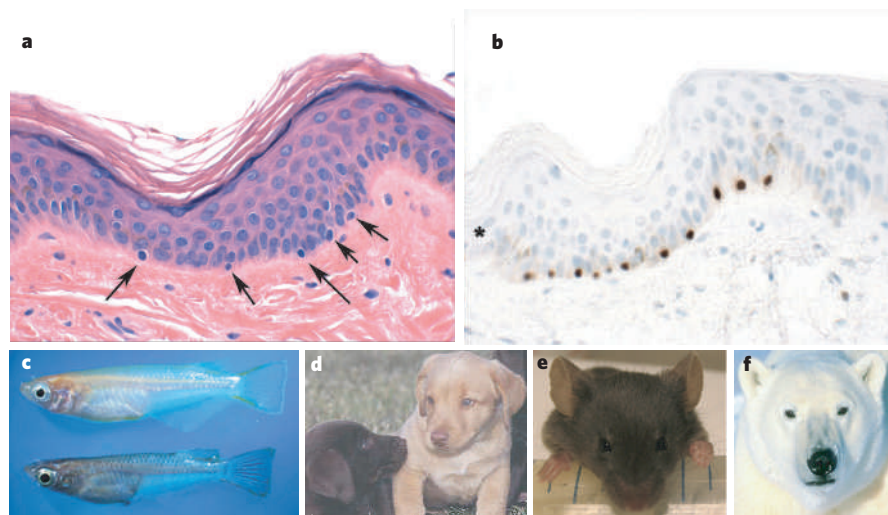
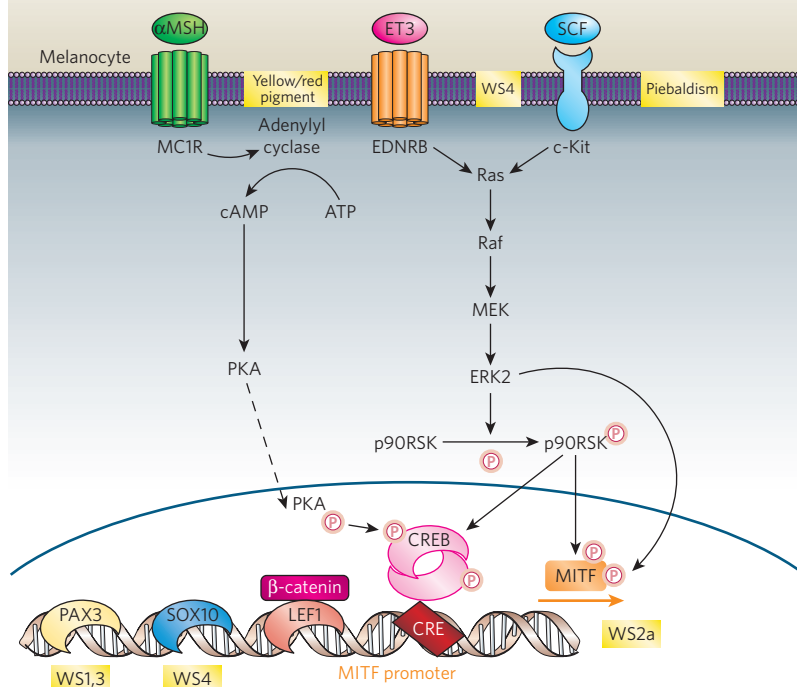


Figure 1 | Vertebrate pigmentation. **a**, Haematoxylin and eosin stain of normal human skin. Cells of the upper layer of the epidermis (keratinocytes) contain large nuclei, which stain blue, and the dermis appears pink as a result of staining of its abundant protein, collagen (muscle and nerve fibres may also stain pink). Normal melanocytes (arrows) have smaller nuclei and inconspicuous cytoplasm compared with the surrounding keratinocytes. Melanocytes are typically located in the basal layer of the epidermis, at the junction with the dermis. Differences in human pigmentation reflect variations in the number of melanosomes in keratinocytes and different melanin granule phenotype (depending on the eumelanin/pheomelanin ratio) rather than variation in the number of melanocytes. (Image courtesy of S. R. Granter, Brigham and Women's Hospital, Massachusetts.) **b**, Immunohistochemical analysis of the same human tissue as shown in **a** identifies melanocytes by using the immunohistochemical marker D5, which stains the MITF transcription factor located in their nuclei (Image courtesy of S. R. Granter.)

c, Medaka goldfish are valuable in pigmentation studies. The wild type, *B/B*, is the lower of the two. Compared with this, *b/b* (top), which is bred for its golden colour, lacks melanin except in the eyes. The *b* locus is highly homologous with the locus for oculocutaneous albinism 4 (OCA-4) in humans, MATP or AIM1. MATP seems to be involved in melanocyte differentiation and melanosome formation. (Image reprinted, with permission, from ref. 82.) **d**, The yellow pigmentation in golden labradors is recessively inherited and results from an amino-acid substitution in *Mc1r* that produces a premature stop codon⁶. The same pigmentation can be seen in mice (*Mc1r*^{cl/c}), horses and cats with hypomorphic *Mc1r* variants. (Image courtesy of Terra Nova.) **e**, Furred animals such as mice lack epidermal melanocytes (except in non-hair-bearing sites such as the ear, nose and paws). **f**, Polar bears have hollow unpigmented fur to blend in with the environment but, unlike other furred animals, have a high density of epidermal melanocytes, which aid in heat retention and produce black skin most notable in non-furred areas. (Image courtesy of First People.)

Figure 2 | The MITF promoter. The MITF promoter is regulated in part by the transcription factors PAX3, SOX10, LEF-1/TCF and CREB during melanocyte development. In humans, mutations affecting the MITF pathway lead to pigmentary and auditory defects that are known collectively as Waardenburg syndrome (the phenotypes of which are highlighted in yellow boxes). The core features are represented by Waardenburg syndrome 2a (WS2a), which results from *MITF* mutations. People with WS1 have craniofacial deformities and those with WS3 can have limb deformities. These are due to different mutations in *PAX3*, a neural-crest-associated transcription factor important for several lineages. WS4 has the additional feature of Hirschsprung's syndrome (megacolon) owing to deficient enteric innervation. Mutations in *SOX10*, endothelin 3 (*ET3*) and its receptor, *EDNRB*, have all been implicated in WS4 and are related to MITF by the pathways illustrated. LEF/TCF is downstream of the WNT/ β -catenin pathway, a key regulator of melanocyte development. MITF transcription is modified by at least two other receptor signalling pathways, those of MC1R and c-Kit, a receptor tyrosine kinase. Hypomorphic/variant alleles of *MC1R* lead to increased pheomelanin/eumelanin ratios, causing blonde/red hair and fair skin. SCF/c-Kit signalling modifies MITF post-translationally through phosphorylation of Ser 73 by the mitogen-activated protein kinase (Ras/Raf/MEK/ERK) pathway, and Ser 409 through the 90 kDa ribosomal S6 protein kinase (p90RSK), thereby altering MITF stability and p300 recruitment (not shown). c-Kit mutations are found in human piebaldism, which presents with extensive skin melanocyte loss and has also been associated with sensorineural deafness, suggesting that piebaldism and WS are in the same spectrum of disorders. PKA, protein kinase A.



Melanocytes and physiological responses

Ultraviolet-induced pigmentation

The most common example of acquired pigmentation is tanning. To the naked eye, the effects of UVR are marked by 'sunburn' and/or 'sun-tan'. Sun sensitivity is the degree of cutaneous inflammation (erythema) and pigmentation that results from exposure to UVR, which can manifest clinically in several ways. Even at low doses of UVR, DNA damage can occur before there is any evidence of change on the skin²⁶.

The tanning response is one of the most striking examples of environmental adaptation in humans. It is well known that α -MSH increases skin darkening in humans, a phenomenon traditionally observed in patients with adrenal insufficiency, whose pituitary glands produce compensatory excesses. Pro-opiomelanocortin (POMC) is the precursor for both α -MSH and ACTH, as well as for other bioactive peptides, including β -endorphin. Although originally identified in the pituitary, POMC production is now known to occur in skin and hair follicles as well^{27,28}. After its production from POMC, α -MSH is secreted by both keratinocytes and melanocytes²⁹. In humans, mutations in *POMC* result in a red-haired phenotype (like that of *MC1R* alleles), as well as metabolic abnormalities such as adrenal insufficiency and obesity³⁰.

There is evidence that DNA damage in itself might be important to the triggering of pigment production. UVR — typically UVB (wavelengths 290–320 nm) — induces thymidine breaks in DNA and the most common signature is cyclobutane dimer formation. Treatment with T4 endonuclease V, an enzyme known to mediate the first and rate-limiting step in repairing thymidine dimers, enhances the tanning response in UV-irradiated melanocytes *in vitro*³¹. Topical application of small DNA fragments such as thymine dinucleotides (pTpT) have been pioneered by Gilchrist and colleagues, and have been shown to induce tyrosinase upregulation and increased pigmentation³¹. The molecular mechanism might be linked to p53, p21 or PCNA (proliferating cell nuclear antigen), levels of which are increased after exposure to pTpT in a time frame similar to that seen after UV irradiation³².

p38 stress-activated kinase has been suggested to participate in UVR-related pigmentation by phosphorylating upstream transcription factor (USF-1), a basic helix–loop–helix leucine zipper transcription factor

that can bind to the tyrosine promoter. In melanocytes derived from *Usf-1*^{−/−} mice, defective UV activation of the POMC and *MC1R* promoters was observed³³. α -MSH has also been shown to activate p38 mitogen-activated protein kinase *in vitro*³⁴. Other pathways implicated include those involving endothelin-1, β -fibroblast growth factor (β -FGF), nitric oxide, p-locus and stem-cell factor (SCF)³⁵. Separate physiological processes — such as X-ray irradiation and DNA-damaging chemotherapeutic agents — can also stimulate a tanning response, which potentially involves pathways that overlap with the UV response³⁶.

Inability to tan highlights several genetic features that might be instructive with regard to the UV-pigmentation response. Variants of *MC1R*, which produce a weakly functioning receptor either through decreased ligand binding or decreased activation of adenylyl cyclase, exhibit relative inability to tan (for example, red-heads cannot tan). Such individuals tend to freckle but do not produce an even, protective pigment. They are also prone to sunburn, although the triggers for this process remain unclear. 'Sunburn cells' — keratinocytes in which apoptosis is triggered by UVR — have been shown by Brash and colleagues to require p53 for their formation³⁷, thus identifying a key role for p53 in modulating a physiological response to a common environmental exposure (UVR), and implicating p53 as a 'guardian of the tissue'³⁸.

Further insight into the UV-tanning response was recently obtained through use of the K14-SCF transgenic mouse¹⁰. By using this transgenic background, fair-skinned mice (*Mc1r*^{+/e}) containing epidermal melanocytes were obtained, and found to be acutely UVR-sensitive. UVR was observed to trigger a more than 30-fold increase in the induction of POMC/ α -MSH in keratinocytes of both mice and humans, suggesting that keratinocytes might have an important role in 'perceiving' UVR and then synthesizing and secreting α -MSH. Mutant *MC1R* was, not surprisingly, seen to ablate any detectable tanning response. However, a rescue strategy of topically delivered forskolin was used to bypass the *Mc1r* mutation and directly activate adenylyl cyclase, and was observed to induce profound skin darkening in genetically fair-skinned mice¹⁰. The induced pigmentation pattern exhibited normal histological features, such as nuclear 'capping' in keratinocytes, and produced significant protection against sunburn cell formation, pyrimidine dimer formation and

skin cancers after UVR exposure¹⁰. These results suggest that the dark pigmentation machinery remains available if appropriately stimulated, and ongoing studies are examining whether this is the case in humans.

Response to UVR and the risk for skin cancer

How does pigmentation protect skin? Although this is presumed to involve straightforward shielding by melanin, our understanding of the process is far from complete. It is clear that skin pigmentation as well as the capacity to tan portend diminished skin cancer risk¹. When the protective effect of melanin has been calculated using minimal erythematous dose measurements, protection for even the darkest-skinned of individuals is no more than 10–15-fold that seen in the absence of melanin (presumably signifying a relatively weak sun-protection factor, SPF (Box 1)). But the factor of protection in terms of skin cancer risk is 500–1,000 (refs 39, 40), indicating that highly pigmented skin is profoundly protected from carcinogenesis. This discrepancy, although subject to numerous caveats, including quantification estimates and endpoint surrogates, suggests that pigment's protective mechanisms might vary for different endpoints such as sunburn and skin cancer.

One possible explanation for the high cancer protection afforded by dark pigmentation might involve mechanism(s) of increased risk associated with blonde/red pigments. *MC1R* variants have been shown to confer an increased risk of melanoma and non-melanoma skin cancers, independently of skin pigment (including red-haired phenotype)^{41,42}. Pheomelanin might function as a photosensitizer capable of generating reactive oxygen species upon UV irradiation⁴³ and has been associated with higher rates of TUNEL (TdT-mediated dUTP nick end labelling)-positive (apoptotic) cells after UV irradiation⁴⁴. Thus, increased pheomelanin production might be a risk factor for melanoma, although the precise underlying mechanism for this process remains to be fully

elucidated to ensure that it is not merely a marker of melanoma risk but is directly involved in carcinogenesis. It should also be noted that whether and how sunscreens protect against cancer is a complex and controversial issue (Box 1).

Melanocyte development

The migratory pathway of the melanoblast

Melanocyte development from its precursor, neural-crest cells, highlights the unique properties of this cell type. Neural-crest cells are pluripotent cells that arise from the dorsalmost point of the neural tube between the surface ectoderm and the neural plate. In addition to melanocytes, neural-crest cells give rise to neurons and glial cells, adrenal medulla, cardiac cells and craniofacial tissue (see ref. 45 for a review). Melanoblasts, the precursors of melanocytes, migrate, proliferate and differentiate en route to their eventual destinations in the basal epidermis and hair follicles, although precise distribution of melanocytes varies among species (Fig. 1e, f).

Melanocyte development has been well characterized in several species, including the mouse embryo. Melanoblast visualization in mice has been achieved by using a *lacZ* transgene in the *Trp2* promoter developed by Jackson and colleagues⁴⁶ (Fig. 3). In mice, melanoblasts differentiate from pluripotent neural-crest cells at about embryonic day (E)8.5, migrating along the dorsolateral pathway and eventually diving ventrally through the dermis. Defects in melanocyte migration typically appear most prominently on the ventral surface as 'white spotting', as this is at the greatest distance (watershed zone) from the dorsum. By E14.5 in mice, melanocytes exit from the overlying dermis and populate the epidermis and developing hair follicles. Melanocytes also reach the choroid of the back of the eye, the iris, the leptomeninges and the stria vascularis of the cochlea (inner ear).

Box 1 | Questions about the benefits of sunscreen

Many people use sunscreens to protect themselves and their children against the most lethal effects of UVR. Unsettlingly, data so far have failed to show that the use of sunscreen protects against melanoma, the deadliest form of skin cancer (see page 851). This and associated controversies have produced responses ranging from deep epidemiological curiosity to a recently filed class-action lawsuit against the sunscreen industry⁸³.

Sunscreens are defined according to 'sun protection factor' (SPF), which is measured by calculating the minimal dose of UVR necessary to cause confluent redness at 24 h after exposure on protected skin compared with unprotected skin. At present, the SPF measurement is based mainly on protection against UVB radiation (wavelengths 290–320 nm), although newer sunscreens may also shield UVA radiation (wavelengths 320–400 nm). UVB can cause DNA damage, and there is growing evidence that UVA might also have carcinogenic effects⁸⁴. Another question raised is whether sunscreen inhibits vitamin D production⁸⁵. However, there is little evidence to suggest that sunscreen prevents adequate vitamin D production⁸⁶, or that low vitamin D levels are associated with increased melanoma risk⁸⁷.

For keratinocyte-derived skin cancers such as squamous-cell carcinoma, the association between UVR and carcinogenesis has been clearly established^{37,88}. Correspondingly, sunscreens, when applied correctly, are effective at reducing the incidence of squamous-cell

carcinoma and its precursor lesion, actinic keratosis⁸⁹. Rates of basal-cell carcinoma, the most common skin cancer, are not as directly affected by sunscreen use⁹⁰. Both types of tumour can be removed surgically, and, although they rarely metastasize, can cause significant morbidity and occasional mortality.

There is even less compelling evidence for sunscreen protection in melanoma. A randomized controlled trial demonstrated decreased naevi (moles) numbers in children using sunscreen⁹¹. The number of naevi present during childhood is a predictor for melanoma⁹², although only a small minority of melanomas have been seen to arise within naevi^{93,94}. At best, naevi counts serve as an indirect measurement. However, the results of several large studies have suggested that sunscreen use is associated with an increased risk of melanoma^{95,96}. Caveats to this conclusion include potential selection bias (higher-risk patients who burn easily are more likely to use sunscreen), questions of proper sunscreen use, and not accounting for changes that have occurred in sunscreen formulation since the studies were performed, including higher UVB and broader UVA protection. This controversy has prompted large meta-analyses^{97,98}, which have not suggested increased melanoma risk with sunscreen use but have also demonstrated no protective effect of sunscreen against melanoma.

Is it mechanistically plausible that sunscreen might not protect against melanoma? At a simplistic level, if the ability to tan easily

correlates with diminished melanoma risk, then sunblocks might convert easy-tanning individuals from low-risk (darkly tanned) to higher-risk (we do not know whether it is the ability to tan or actually being tanned that confers true protection). A remarkable feature of melanoma risk is the identification of *MC1R* as a melanoma risk factor gene, independently of UVR exposure^{41,99}. Blonde/red pheomelanin pigments have been suggested to enhance intrinsic DNA damage in cells, particularly in response to UVR^{42,100}. Although this observation requires more systematic epidemiological study, it is possible that pheomelanin is weakly carcinogenic and actually contributes to melanoma formation.

We cannot assume that sunscreens protect against melanoma analogously to their protection against sunburn (which is demonstrated by 'missed spots' during application). It remains possible that the kinetics of sunburn, squamous cell carcinoma, and melanoma formation are so different that sunscreen protection from melanoma requires longer follow-up than studies so far have reported. Alternative dosing/applications of sunscreens might offer different protection against these endpoints. Importantly, the data so far do not suggest that sunscreens be abandoned. On the contrary, protection from sunburn and squamous cell carcinoma are important. But protection against melanoma is more complex and an alternative prevention strategy for now is to stay out of the sun.

Various signalling pathways and transcription factors tightly regulate melanocyte migration. These proteins and pathways provide and integrate spatial and temporal signals to create the proper environment for normal development and migration. Mutations in genes affecting this process produce hypopigmentation that arises from lack of melanocytes rather than lack of pigment in viable melanocytes, as occurs in albinism. Key genes in this developmental pathway include *PAX3* (paired-box 3), *SOX10* (sex-determining region Y (SRY)-box 10), *MITF*, endothelin 3 and endothelin receptor B (*EDNRB*). Disruption of these genes has led to clearer understanding of certain human inherited pigmentation disorders, specifically Waardenburg Syndrome (WS), which is characterized by hearing and pigmentary defects. The highly related Tietz syndrome has a similar phenotype but is associated with dominant-negative mutations in *MITF*. Given the defined molecular and phenotypic relationships among the genes responsible for these syndromes, they represent an epistatic tree of interacting melanocytic regulatory factors⁴⁷ (Fig. 2).

Microphthalmia transcription factor

The core phenotype of Waardenburg Syndrome is seen in WS type 2 and is characterized by congenital white forelock, asymmetric iris colour (heterochromia irides) and sensorineural deafness, all resulting from disrupted melanocyte migration. A subset of WS type 2 patients (WS2a) have germline heterozygous mutations in *MITF*. *Mitf* was originally described as a mouse coat colour mutant more than 60 years ago. Homozygous mutant mice have small eyes, white fur and deafness. This is representative of the affected cell lineages, which include retinal pigment epithelial (RPE) cells (*Mitf* mutations result in microphthalmia), osteoclasts (osteopetrosis is seen with certain mutant alleles in homozygotes) and mast cells (for which mutation causes severe dysfunction)⁴⁸. Numerous *Mitf* mutations have been identified in mice as well as other species, and have provided considerable insight into this gene locus and its central role in melanocyte development⁴⁸.

MITF is a member of the Myc-related family of basic helix–loop–helix leucine zipper (bHLH–Zip) transcription factors and is conserved in essentially all vertebrate species⁴⁸. Like other bHLH–Zip factors, it binds to the canonical (CA[T/C]GTG) E box sequence⁴⁸. The MiT subfamily of bHLH–Zip factors includes *MITF*, *TFEB*, *TFE3* and *TFEC*⁴⁸. These MiT factors homodimerize and heterodimerize in all combinations, although restricted tissue expression limits the availability of different partner combinations in specific contexts. In melanocytes, *MITF* alone seems to be the critical family member, whereas *MITF* and *TFE3* heterodimerize and exhibit functional redundancy in osteoclasts⁴⁹. Three MiT factors, including *MITF*, have also been identified as human oncogenes and found to be involved in multiple malignancies, including melanoma⁵⁰ (see page 851).

The WNT/ β -catenin–signalling pathway is also essential for neural crest induction and melanocyte development. Mice lacking wild-type *Wnt1* and *Wnt3a* have pigmentation defects⁵¹. WNT1 and WNT3A trigger a canonical pathway resulting in β -catenin–induced transcription at TCF/LEF (T-cell factor/lymphoid enhancer factor) promoter/enhancer elements (Fig. 2). Numerous targets of TCF/LEF have been identified, including *Myc* and cyclin D1 in non-melanocytes, and *MITF*, *TRP2* and *SOX10* in melanocytes and melanoma cells (see ref. 52 for a review). Overexpression of β -catenin in zebrafish promotes melanoblast formation and reduces the formation of neurons and glia⁵³.

Melanocyte homing to the epidermis and hair follicles

c-Kit is a tyrosine kinase receptor involved in melanoblast expansion, survival and migration⁵⁴. Activation of *c-Kit* by Kit-ligand (KitL, also known as steel factor or SCF) leads to Ras activation and multiple canonical signalling as well as post-translational modification of *MITF*⁵⁵ (Fig. 2). *c-Kit*, *SCF* and *SNAIL2* (also known as *SLUG*) mutations have all been identified in human piebaldism, an autosomal dominant ventral depigmentary disorder^{56,57}. Piebaldism is characterized by white forelock but, unlike WS, is not accompanied by deafness and more commonly results in white, depigmented (melanocyte-free) areas of

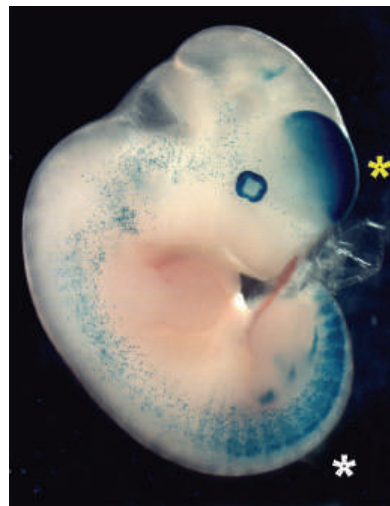


Figure 3 | Labelled melanoblasts in *DCT-lacZ* mouse embryos. The *DCT-lacZ* transgenic mouse labels melanoblasts and melanocytes (blue). On E11.5, labelled melanoblasts can be seen at high density flanking the neural tube (white asterisk) and in the cervical region (yellow asterisk). After dorsolateral migration, the melanoblasts begin to migrate ventrally. (Image reprinted, with permission, from ref. 46.)

skin rather than hair involvement. However, there have been reports of a *c-Kit* mutation carrier with sensorineural deafness and no cutaneous pigmentary changes, suggesting a potential overlap syndrome between piebaldism and WS⁵⁸.

The availability of SCF has a crucial role in permitting melanoblast survival and promoting proliferation, both initially in the dorsolateral migration pathway and later from the dermal mesenchyme to colonize the hair follicles and epidermis^{46,59,60}. Expression of transgenic *Scf* under a keratin promoter (K14), is sufficient to support melanocyte homing to the epidermis in furry animals such as mice, which otherwise have few epidermal melanocytes in fur-bearing regions⁴⁶ (Fig. 1e, f). The use of imatinib mesylate (also known as STI 571 or Gleevec), a potent BCR–ABL tyrosine kinase inhibitor that also inhibits *c-Kit* tyrosine kinase, can result in loss of pigment (loss of melanocytes) in the skin⁶¹.

SLUG encodes a zinc finger transcription factor associated with piebaldism. It was first noted in mice that *Slug* mutations caused a phenotype very similar to piebaldism with anaemia, infertility and white forelock, and depigmentation of the ventral trunk, tail and feet⁶². Humans with piebaldism lacking *c-Kit* mutations were found to have heterozygous deletions encompassing the *SLUG* coding region⁶³. In addition, *SLUG* mutation was also reported in WS2, and a mechanism proposed to account for these effects involves the binding of *MITF* to the *SLUG* promoter⁶³. *SLUG* seems to be required for melanoblast migration and/or survival⁶². In a mouse model of melanoma, the inhibition of *SLUG* by small interfering RNA (siRNA) suppressed metastatic propensity, potentially linking *SLUG* to both migration and metastasis-related behaviours⁶⁴.

In addition to *c-Kit*/SCF, other mechanisms are likely to be involved in the late steps of melanocyte migration from the dermis into the epidermis. These include endothelins 1 and 3, hepatocyte growth factor (HGF) and basic FGF⁶⁵. Evidence for HGF's influence on melanocyte homing has been obtained by Merlino and colleagues, who discovered extensive melanosis at diverse sites in transgenic mice expressing HGF driven by the metallothionein promoter⁶⁶. This model has been particularly interesting because of its high-penetrance melanoma formation after single neonatal doses of UVR⁶⁷. Cadherins are also implicated: as dermal melanoblasts move through the basement membrane they express E-cadherin, which is then downregulated and replaced by P-cadherin during migration into the hair follicles⁶⁸. In mice, a screen of dominant phenotypes by Barsh and colleagues identified two new classes of mutant with dark skin (*Dsk*)⁶⁹. The first group contained activating mutations in G proteins, *Gna11* and *Gnaq*, producing excess melanoblasts in the dermis even though the correct number reached the epidermis and hair follicles⁷⁰. The *Dsk* mutations were able to rescue pigmentary defects from loss of *PAX3* and *c-Kit*, but did not do so in *Ednrb*^{−/−} mice⁷¹, suggesting a cell-autonomous amplification of normal endothelin signalling through which mutant G_q subunits cause an excess number of early melanoblasts.

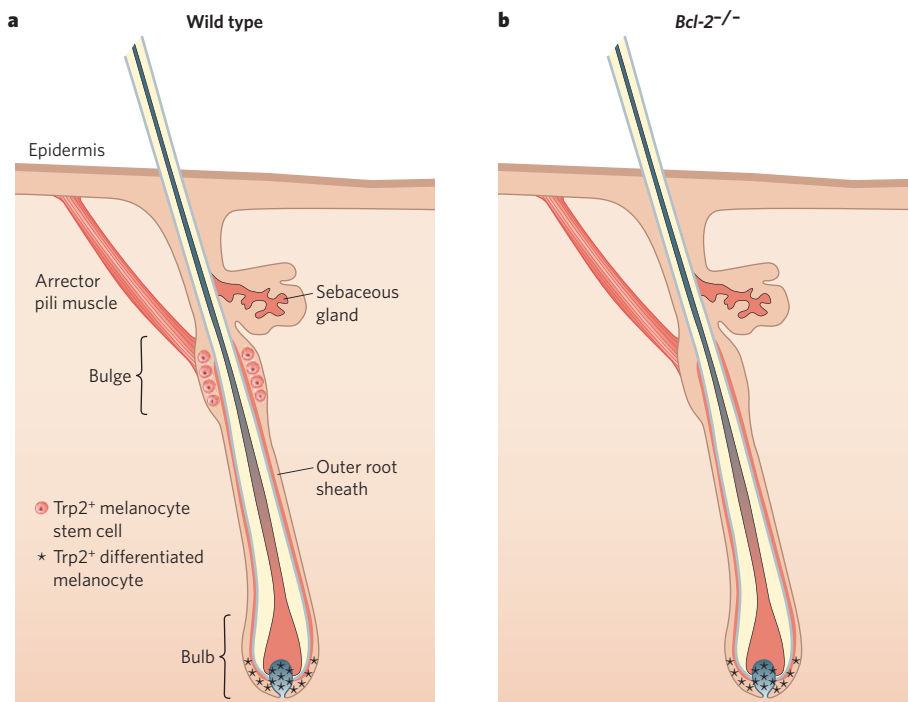


Figure 4 | *DCT-lacZ* melanoblasts in mouse hair follicles. Using *Dct(Trp2)-lacZ* transgenic mice, melanocyte stem cells can be revealed in the bulge region of a hair follicle, geographically distinct from the differentiated melanocytes in the hair bulb, as demonstrated in this diagram. As part of the hair follicle cycle, differentiated melanocytes migrate from the bulge to the bulb, where they export pigment to hair-producing keratinocytes. In *Bcl-2*^{-/-} mice, normal melanocyte stem-cell and differentiated melanocyte populations are seen at birth. At postnatal day 8.5, the bulge melanocytes abruptly disappear, whereas in controls (*Bcl-2*^{+/+}) and bulb melanocytes of the same *Bcl-2*^{-/-} follicles these cells survive until the next catagen/regressive follicle stage. Subsequent hair greying in *Bcl-2*^{-/-} mice is thought to arise from lack of regenerative melanocyte stem cells.

Melanoblast survival

MITF expression is activated early on during the transition from pluripotent neural-crest cells to melanoblasts and is required for the survival of migrating melanoblasts⁴⁸. Complete deficiency of MITF results in absence of melanocytes, suggesting that MITF is essential for lineage survival, for proliferation or to prevent transdifferentiation towards other neural-crest lineages (such as glia and neurons). Evidence that MITF remains vital for melanocyte survival throughout the life of an organism comes from the hypomorphic mutant allele *Mitf*^{nit}, which exhibits near-normal melanocyte development, but accelerated age-dependent greying of coat colour due to postnatal melanocyte loss⁷².

Bcl-2 (B-cell leukaemia/lymphoma 2), a transcriptional target of MITF⁷³ and known inhibitor of apoptosis, is also required for melanocyte viability⁷⁴. Cyclin-dependent kinase 2 (Cdk2), a cell-cycle regulator, is another MITF target⁷⁵. Melanoma cells have been shown to require Cdk2 expression to maintain their cell cycle and viability⁷⁵. Nevertheless, MITF's interactions with *Bcl-2* and Cdk2 do not fully explain melanocytes' requirement for MITF to remain viable during development, because the *Mitf*-null homozygous phenotype is more severe than either of the single knockouts, and expression of either gene in *Mitf*-knocked-down melanoma cells only partly rescues viability^{75,76}.

The cytokine receptor c-Met has recently been found to be another direct transcriptional target of MITF⁷⁶. Many additional targets of MITF have been identified, including the CDK inhibitors p16^{INK4a} and p21/Cip1 (see ref. 9 for a review). Together, they probably contribute to MITF's dual activities in relation to melanocytic differentiation and survival/proliferation, and some may serve as surrogate drug targets for the MITF oncogenic pathway. The function of MITF-mediated survival should be distinguished from differentiation markers, as MITF is believed to coordinately regulate the expression of pigment genes, although probably in conjunction with other context-dependent factors. In addition to tyrosinase, TRP1 and TRP2, pigment targets of MITF in melanocytes include silver (*PMEL17*, also known as gp100, which encodes the melanoma-diagnostic epitope HMB45), melan-A (also known as MART1), melastatin (TRPM1) and AIM1 (oculocutaneous albinism 4)⁹.

Melanocyte stem cells and grey hair

Melanocytes of the hair follicle are responsible for hair pigment. The life-cycle of the hair follicle melanocyte is closely linked to that of the rest of the hair follicle, which is typically in growth (anagen), but moves through

a brief period of regression (catagen) and finally enters a resting phase (telogen) in which the hair is released and the cycle can begin anew (see page 834). Multipotent epidermal stem cells exist in the bulge region (at the bottom of the permanent portion of the follicle, just below the sebaceous gland)⁷⁷ (Fig. 4). This is also where melanocyte stem cells reside. These are responsible for restoring the pool of differentiated melanocytes to the hair bulb, where the pigment is incorporated while new hair is being synthesized⁴⁶ (Fig. 3). The key features satisfying the definition of 'stemness' include slow cycling, self-maintenance, immaturity and competence to regenerate progeny when appropriately stimulated⁷⁸. Independent validation that the melanocyte stem-cell niche resides in hair follicles comes from patients with vitiligo (absence of epidermal melanocytes) resulting from immunosuppressive treatment, in whom repigmentation is clearly initiated in radial fashion from hair follicles.

The *Dct-lacZ* transgenic mouse⁴⁶ has helped researchers to identify the melanoblast population in the bulge region (Fig. 3). The power of this system was illustrated by the analysis of several mouse models of hair greying⁷⁸. *Bcl-2*^{-/-} mice are born with normal pigment but turn grey at the second hair follicle cycle⁷⁴. At birth, *Bcl-2*^{-/-} mice exhibit normal melanocyte morphology and differentiation, although it has been suggested that melanocyte stem-cell formation could be deficient, at least in certain follicles⁷⁹. Between postnatal days 6 and 8 (ref. 80) there is an abrupt loss of all bulge melanocytes (Fig. 4), whereas bulb melanocytes remain until the end of the first follicle cycle⁷⁸. This striking phenotype has been interpreted to suggest a selective requirement for *Bcl-2* in the maintenance of melanocyte stem cells at a crucial stage in the anagen phase of the follicle cycle⁷⁸.

More gradual greying has been observed in the hypomorphic mouse mutant *Mitf*^{nit} (ref. 78). In this case, bulge melanocytes were lost more gradually, although still in a markedly accelerated manner compared with the wild type. Loss of the stem-cell population was preceded by the appearance of unexpected pigment-producing cells in the bulge region (stem-cell niche). Because differentiation/pigmentation is generally incompatible with stemness it has been proposed that this 'inappropriate' bulge pigmentation might signify aberrant exit from the melanocyte stem-cell pool. Eventual depletion of stem cells in this way would produce grey/white hairs in the subsequent follicle cycle.

Age-related greying in both mouse and human follicles is also accompanied by bulge melanocyte attrition and is preceded by 'inappropriate pigmentation' in that population of (niche) melanocytes⁷⁸. But it remains

to be seen whether other causes of human greying exist, such as in cases that are followed by unexpected darkening (as occasionally reported in patients using imatinib⁸¹), which presumably do not involve depletion of melanocyte stem cells. Such instances might be more amenable to 'rescue' strategies for greyness.

Future prospects

Melanocyte biology remains at the crossroads of the laboratory and the clinic. Numerous pigmentation mutants are phenotypically profound but remain mechanistically uncharacterized. Although pigment mutants are informative with regard to basic genetic and developmental mechanisms, they also provide important clues to processes of photoprotection, cancer predisposition and even human evolution. The elucidation of genetic regulators of pigmentation will probably increase even more markedly with the advent of targeted knockouts that spare embryonic lethality for vital genes.

Certain lineage-specific features of melanocytes have attracted well-deserved attention and leave important questions to be answered. The dispensability of melanocytes to hair follicle viability offers the opportunity to identify and discover genes of particular importance to the melanocyte lineage. The physical separation of melanocyte stem cells from their differentiated progeny also offers distinct experimental advantages. Key questions that remain to be answered include what the signal(s) are that modulate melanocyte stem cells' awakening from and return to quiescence. It will be important to determine the signals modulating stem-cell attrition and 'inappropriate pigmentation' of melanocytes in the bulge that precede stem-cell loss and hair greying. To what degree is the melanocyte stem cell dependent on the specialized niche, and to what degree might it be harnessed for other uses, such as repigmentation of scars or artificial skin? What are the signals modulating melanocyte death during catagen? These questions also stand to provide clues relevant to the biology of epidermal melanocytes, whose intercellular communication is also of enormous importance but only partly understood.

Dysregulation of melanocytic growth or survival is an enormously common occurrence. Almost all people harbour multiple moles, which are benign, senescent neoplasias that only rarely (if ever) transform to invasive melanomas (see page 851). Are BRAF^{V600E}-transformed melanocytes the most common instance of such clonal transformation, or do similar benign neoplasms occur in most tissues, but remain unapparent owing to absence of pigment? Why do melanocytic neoplasms so uncommonly harbour UVR-signature mutations, especially when melanoma incidence is related to the combination of skin phototype and UVR exposure? Could it be that keratinocytes are the primary UV-respondering population, and melanoma formation is largely a consequence of chronic reactive secondary stimulation/proliferation (akin to other chronic-inflammation-associated cancers)?

Skin is the most common site of cancer in man, yet its malignancies are likely to be among the most preventable. A considerable fraction are caused by the known environmental carcinogen, UVR. Behaviours including specific sun-seeking activities abound in many of today's societies. Sun exposure in other cases is casual and unintentional, whereas in the case of sunscreen may be associated with misconceptions involving the types of protection provided. It is likely that opportunities to improve our understanding of melanocytic contributions to cancer will arise increasingly. An ability to exploit melanin and its photoprotective effects might ameliorate cancer risk in vulnerable populations. The availability of ever-improving animal models, high-quality human diagnostic reagents, rich data resources from the post-genomic era and increasing opportunities for targeted therapeutics will hopefully permit new means of safely modulating the melanocyte and its activities for purposes ranging from the development of cosmetics to the prevention of cancer. ■

1. Fitzpatrick, T. B. The validity and practicality of sun-reactive skin types I through VI. *Arch. Dermatol.* **124**, 869–871 (1988).
2. Ziegler, I. The pteridine pathway in zebrafish: regulation and specification during the determination of neural crest cell-fate. *Pigment Cell Res.* **16**, 172–182 (2003).
3. Rees, J. L. Genetics of hair and skin color. *Annu. Rev. Genet.* **37**, 67–90 (2003).
4. Searle, A. G. An extension series in the mouse. *J. Hered.* **59**, 341–342 (1968).

5. Robbins, L. S. *et al.* Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* **72**, 827–834 (1993).
6. Everts, R. E., Rothuizen, J. & van Oost, B. A. Identification of a premature stop codon in the melanocyte-stimulating hormone receptor gene (MC1R) in labrador and golden retrievers with yellow coat colour. *Anim. Genet.* **31**, 194–199 (2000).
7. Balthasar, N. *et al.* Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell* **123**, 493–505 (2005).
8. Mountjoy, K. G., Robbins, L. S., Mortrud, M. T. & Cone, R. D. The cloning of a family of genes that encode the melanocortin receptors. *Science* **257**, 1248–1251 (1992).
9. Levy, C., Khaled, M. & Fisher, D. E. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol. Med.* **12**, 406–414 (2006).
10. D'Orazio, J. A. *et al.* Topical drug rescue strategy and skin protection based on the role of MC1R in UV-induced tanning. *Nature* **443**, 340–344 (2006).
11. Barsh, G., Gunn, T., He, L., Schlossman, S. & Duke-Cohan, J. Biochemical and genetic studies of pigment-type switching. *Pigment Cell Res.* **13** (Suppl. 8), 48–53 (2000).
12. Furumura, M., Sakai, C., Abdel-Malek, Z., Barsh, G. S. & Hearing, V. J. The interaction of agouti signal protein and melanocyte stimulating hormone to regulate melanin formation in mammals. *Pigment Cell Res.* **9**, 191–203 (1996).
13. Kanetsky, P. A. *et al.* A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *Am. J. Hum. Genet.* **70**, 770–775 (2002).
14. Rana, B. K. *et al.* High polymorphism at the human melanocortin 1 receptor locus. *Genetics* **151**, 1547–1557 (1999).
15. Healy, E. *et al.* Functional variation of MC1R alleles from red-haired individuals. *Hum. Mol. Genet.* **10**, 2397–2402 (2001).
16. Ringholm, A. *et al.* Pharmacological characterization of loss of function mutations of the human melanocortin 1 receptor that are associated with red hair. *J. Invest. Dermatol.* **123**, 917–923 (2004).
17. Naysmith, L. *et al.* Quantitative measures of the effect of the melanocortin 1 receptor on human pigmentary status. *J. Invest. Dermatol.* **122**, 423–428 (2004).
18. Harding, R. M. *et al.* Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **66**, 1351–1361 (2000).
19. Mogil, J. S. *et al.* The melanocortin-1 receptor gene mediates female-specific mechanisms of analgesia in mice and humans. *Proc. Natl Acad. Sci. USA* **100**, 4867–4872 (2003).
20. Lamason, R. L. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005).
21. Slominski, A., Tobin, D. J., Shibahara, S. & Wortsman, J. Melanin pigmentation in mammalian skin and its hormonal regulation. *Physiol. Rev.* **84**, 1155–1228 (2004).
22. Land, E. J. & Riley, P. A. Spontaneous redox reactions of dopaquinone and the balance between the eumelanin and pheomelanin pathways. *Pigment Cell Res.* **13**, 273–277 (2000).
23. Oetting, W. S., Fryer, J. P., Shriram, S. & King, R. A. Oculocutaneous albinism type 1: the last 100 years. *Pigment Cell Res.* **16**, 307–311 (2003).
24. Wei, M. L. Hermansky-Pudlak syndrome: a disease of protein trafficking and organelle function. *Pigment Cell Res.* **19**, 19–42 (2006).
25. Boissy, R. E. Melanosome transfer to and translocation in the keratinocyte. *Exp. Dermatol.* **12** (Suppl. 2), 5–12 (2003).
26. Heenen, M., Giacconi, P. U. & Golstein, P. Individual variations in the correlation between erythral threshold, UV-induced DNA damage and sun-burn cell formation. *J. Photochem. Photobiol. B* **63**, 84–87 (2001).
27. Tasmali, M., Ancans, J., Yukitake, J. & Thody, A. J. Skin POMC peptides: their actions at the human MC-1 receptor and roles in the tanning response. *Pigment Cell Res.* **13** (Suppl. 8), 125–129 (2000).
28. Paus, R. *et al.* The skin POMC system (SPS). Leads and lessons from the hair follicle. *Ann. NY Acad. Sci.* **885**, 350–363 (1999).
29. Schauer, E. *et al.* Proopiomelanocortin-derived peptides are synthesized and released by human keratinocytes. *J. Clin. Invest.* **93**, 2258–2262 (1994).
30. Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature Genet.* **19**, 155–157 (1998).
31. Eller, M. S., Yaar, M. & Gilchrist, B. A. DNA damage and melanogenesis. *Nature* **372**, 413–414 (1994).
32. Eller, M. S., Ostrom, K. & Gilchrist, B. A. DNA damage enhances melanogenesis. *Proc. Natl Acad. Sci. USA* **93**, 1087–1092 (1996).
33. Corre, S. *et al.* UV-induced expression of key component of the tanning process, the POMC and MC1R genes, is dependent on the p-38-activated upstream stimulating factor-1 (USF-1). *J. Biol. Chem.* **279**, 51226–51233 (2004).
34. Smalley, K. & Eisen, T. The involvement of p38 mitogen-activated protein kinase in the alpha-melanocyte stimulating hormone (alpha-MSH)-induced melanogenic and anti-proliferative effects in B16 murine melanoma cells. *FEBS Lett.* **476**, 198–202 (2000).
35. Ancans, J., Flanagan, N., Hoogduijn, M. J. & Thody, A. J. P-locus is a target for the melanogenic effects of MC-1R signaling: a possible control point for facultative pigmentation. *Ann. NY Acad. Sci.* **994**, 373–377 (2003).
36. Suzuki, K., Ojima, M., Kodama, S. & Watanabe, M. Radiation-induced DNA damage and delayed induced genomic instability. *Oncogene* **22**, 6988–6993 (2003).
37. Ziegler, A. *et al.* Sunburn and p53 in the onset of skin cancer. *Nature* **372**, 773–776 (1994).
38. Kamb, A. Sun protection factor p53. *Nature* **372**, 730–731 (1994).
39. Kaidbey, K. H., Agin, P. P., Sayre, R. M. & Kligman, A. M. Photoprotection by melanin — a comparison of black and Caucasian skin. *J. Am. Acad. Dermatol.* **1**, 249–260 (1979).
40. Kollias, N., Sayre, R. M., Zeise, L. & Chedekel, M. R. Photoprotection by melanin. *J. Photochem. Photobiol. B* **9**, 135–160 (1991).
41. Kennedy, C. *et al.* Melanocortin 1 receptor (MC1R) gene variants are associated with an increased risk for cutaneous melanoma which is largely independent of skin type and hair color. *J. Invest. Dermatol.* **117**, 294–300 (2001).
42. Scott, M. C. *et al.* Human melanocortin 1 receptor variants, receptor function and melanocyte response to UV radiation. *J. Cell Sci.* **115**, 2349–2355 (2002).
43. Hill, H. Z. & Hill, G. J. UVA, pheomelanin and the carcinogenesis of melanoma. *Pigment Cell Res.* **13** (Suppl. 8), 140–144 (2000).
44. Takeuchi, S. *et al.* Melanin acts as a potent UVB photosensitizer to cause an atypical mode of cell death in murine skin. *Proc. Natl Acad. Sci. USA* **101**, 15076–15081 (2004).

45. Le Douarin, N. M., Creuzet, S., Couly, G. & Dupin, E. Neural crest cell plasticity and its limits. *Development* **131**, 4637–4650 (2004).
46. Mackenzie, M. A., Jordan, S. A., Budd, P. S. & Jackson, I. J. Activation of the receptor tyrosine kinase Kit is required for the proliferation of melanoblasts in the mouse embryo. *Dev. Biol.* **192**, 99–107 (1997).
47. Price, E. R. & Fisher, D. E. Sensorineural deafness and pigmentation genes: melanocytes and the Mitf transcriptional network. *Neuron* **30**, 15–18 (2001).
48. Steingrimsson, E., Copeland, N. G. & Jenkins, N. A. Melanocytes and the microphthalmia transcription factor network. *Annu. Rev. Genet.* **38**, 365–411 (2004).
49. Hershey, C. L. & Fisher, D. E. Mitf and Tfe3: members of a b-HLH-ZIP transcription factor family essential for osteoclast development and function. *Bone* **34**, 689–696 (2004).
50. Chin, L., Garraway, L. A. & Fisher, D. E. Malignant melanoma: genetics and therapeutics in the genomic era. *Genes Dev.* **20**, 2149–2182 (2006).
51. Christiansen, J. H., Coles, E. G. & Wilkinson, D. G. Molecular control of neural crest formation, migration and differentiation. *Curr. Opin. Cell Biol.* **12**, 719–724 (2000).
52. Larue, L. & Delmas, V. The WNT/ β -catenin pathway in melanoma. *Front. Biosci.* **11**, 733–742 (2006).
53. Lee, H. O., Levorse, J. M. & Shin, M. K. The endothelin receptor-B is required for the migration of neural crest-derived melanocyte and enteric neuron precursors. *Dev. Biol.* **259**, 162–175 (2003).
54. Wehrle-Haller, B. The role of Kit-ligand in melanocyte development and epidermal homeostasis. *Pigment Cell Res.* **16**, 287–296 (2003).
55. Hemesath, T. J., Price, E. R., Takemoto, C., Badalian, T. & Fisher, D. E. MAP kinase links the transcription factor Microphthalmia to c-Kit signalling in melanocytes. *Nature* **391**, 298–301 (1998).
56. Ruan, H. B., Zhang, N. & Gao, X. Identification of a novel point mutation of mouse proto-oncogene *c-kit* through N-ethyl-N-nitrosourea mutagenesis. *Genetics* **169**, 819–831 (2005).
57. Spritz, R. A. & Beighton, P. Piebaldism with deafness: molecular evidence for an expanded syndrome. *Am. J. Med. Genet.* **75**, 101–103 (1998).
58. Shears, D. et al. Molecular heterogeneity in two families with auditory pigmentary syndromes: the role of neuroimaging and genetic analysis in deafness. *Clin. Genet.* **65**, 384–389 (2004).
59. Steel, K. P., Davidson, D. R. & Jackson, I. J. TRP-2/DT, a new early melanoblast marker, shows that steel growth factor (c-kit ligand) is a survival factor. *Development* **115**, 1111–1119 (1992).
60. Jordan, S. A. & Jackson, I. J. MGF (KIT ligand) is a chemokinetic factor for melanoblast migration into hair follicles. *Dev. Biol.* **225**, 424–436 (2000).
61. Legros, L., Cassuto, J. P. & Ortonne, J. P. Imatinib mesilate (Glivec): a systemic depigmenting agent for extensive vitiligo? *Br. J. Dermatol.* **153**, 691–692 (2005).
62. Perez-Losada, J. et al. Zinc-finger transcription factor Slug contributes to the function of the stem cell factor c-kit signaling pathway. *Blood* **100**, 1274–1286 (2002).
63. Sanchez-Martin, M. et al. Deletion of the SLUG (SNAI2) gene results in human piebaldism. *Am. J. Med. Genet. A* **122**, 125–132 (2003).
64. Gupta, P. B. et al. The melanocyte differentiation program predisposes to metastasis after neoplastic transformation. *Nature Genet.* **37**, 1047–1054 (2005).
65. Imokawa, G. Autocrine and paracrine regulation of melanocytes in human skin and in pigmentary disorders. *Pigment Cell Res.* **17**, 96–110 (2004).
66. Otsuka, T. et al. c-Met autocrine activation induces development of malignant melanoma and acquisition of the metastatic phenotype. *Cancer Res.* **58**, 5157–5167 (1998).
67. Noonan, F. P. et al. Neonatal sunburn and melanoma in mice. *Nature* **413**, 271–272 (2001).
68. Nishimura, E. K., Yoshida, H., Kunisada, T. & Nishikawa, S. I. Regulation of E- and P-cadherin expression correlated with melanocyte migration and diversification. *Dev. Biol.* **215**, 155–166 (1999).
69. Fitch, K. R. et al. Genetics of dark skin in mice. *Genes Dev.* **17**, 214–228 (2003).
70. Van Raamsdonk, C. D., Fitch, K. R., Fuchs, H., de Angelis, M. H. & Barsh, G. S. Effects of G-protein mutations on skin color. *Nature Genet.* **36**, 961–968 (2004).
71. Shin, M. K., Levorse, J. M., Ingram, R. S. & Tilghman, S. M. The temporal requirement for endothelin receptor-B signalling during neural crest development. *Nature* **402**, 496–501 (1999).
72. Lerner, A. B. et al. A mouse model for vitiligo. *J. Invest. Dermatol.* **87**, 299–304 (1986).
73. McGill, G. G. et al. Bcl2 regulation by the melanocyte master regulator Mitf modulates lineage survival and melanoma cell viability. *Cell* **109**, 707–718 (2002).
74. Veis, D. J., Sorenson, C. M., Shutter, J. R. & Korsmeyer, S. J. Bcl-2-deficient mice demonstrate fulminant lymphoid apoptosis, polycystic kidneys, and hypopigmented hair. *Cell* **75**, 229–240 (1993).
75. Du, J. et al. Critical role of CDK7 for melanoma growth linked to its melanocyte-specific transcriptional regulation by MITF. *Cancer Cell* **6**, 565–576 (2004).
76. McGill, G. G., Haq, R., Nishimura, E. K. & Fisher, D. E. c-Met expression is regulated by Mitf in the melanocyte lineage. *J. Biol. Chem.* **281**, 10365–10373 (2006).
77. Moore, K. A. & Lemischka, I. R. Stem cells and their niches. *Science* **311**, 1880–1885 (2006).
78. Nishimura, E. K., Granter, S. R. & Fisher, D. E. Mechanisms of hair graying: incomplete melanocyte stem cell maintenance in the niche. *Science* **307**, 720–724 (2005).
79. Mak, S. S., Moriyama, M., Nishioka, E., Osawa, M. & Nishikawa, S. Indispensable role of Bcl2 in the development of the melanocyte stem cell. *Dev. Biol.* **291**, 144–153 (2006).
80. Paus, R. et al. A comprehensive guide for the recognition and classification of distinct stages of hair follicle morphogenesis. *J. Invest. Dermatol.* **113**, 523–532 (1999).
81. Tsao, A. S., Kantarjian, H., Cortes, J., O'Brien, S. & Talpaz, M. Imatinib mesylate causes hypopigmentation in the skin. *Cancer* **98**, 2483–2487 (2003).
82. Fukamachi, S., Shimada, A. & Shima, A. Mutations in the gene encoding B, a novel transporter protein, reduce melanin content in medaka. *Nature Genet.* **28**, 381–385 (2001).
83. Olson, E. The rub on sunscreen. *New York Times*, 19 June 2006.
84. Wang, S. Q. et al. Ultraviolet A and melanoma: a review. *J. Am. Acad. Dermatol.* **44**, 837–846 (2001).
85. Berwick, M. et al. Sun exposure and mortality from melanoma. *J. Natl Cancer Inst.* **97**, 195–199 (2005).
86. Solitto, R. B., Kraemer, K. H. & DiGiovanna, J. J. Normal vitamin D levels can be maintained despite rigorous photoprotection: six years' experience with xeroderma pigmentosum. *J. Am. Acad. Dermatol.* **37**, 942–947 (1997).
87. Weinstock, M. A., Stampfer, M. J., Lew, R. A., Willett, W. C. & Sober, A. J. Case-control study of melanoma and dietary vitamin D: implications for advocacy of sun protection and sunscreen use. *J. Invest. Dermatol.* **98**, 809–811 (1992).
88. Li, G., Tron, V. & Ho, V. Induction of squamous cell carcinoma in p53-deficient mice after ultraviolet irradiation. *J. Invest. Dermatol.* **110**, 72–75 (1998).
89. Green, A. et al. Daily sunscreen application and betacarotene supplementation in prevention of basal-cell and squamous-cell carcinomas of the skin: a randomised controlled trial. *Lancet* **354**, 723–729 (1999).
90. Vainio, H., Miller, A. B. & Bianchini, F. An international evaluation of the cancer-preventive potential of sunscreens. *Int. J. Cancer* **88**, 838–842 (2000).
91. Gallagher, R. P. et al. Broad-spectrum sunscreen use and the development of new nevi in white children: a randomized controlled trial. *J. Am. Med. Assoc.* **283**, 2955–2960 (2000).
92. Pfahler, A. et al. Monitoring of nevus density in children as a method to detect shifts in melanoma risk in the population. *Prev. Med.* **38**, 382–387 (2004).
93. Dennis, L. K. et al. Constitutional factors and sun exposure in relation to nevi: a population-based cross-sectional study. *Am. J. Epidemiol.* **143**, 248–256 (1996).
94. Mones, J. M. & Ackerman, A. B. Melanomas in prepubescent children: review comprehensively, critique historically, criteria diagnostically, and course biologically. *Am. J. Dermatopathol.* **25**, 223–238 (2003).
95. Wolf, P., Quehenberger, F., Mullegger, R., Stranz, B. & Kerl, H. Phenotypic markers, sunlight-related factors and sunscreen use in patients with cutaneous melanoma: an Austrian case-control study. *Melanoma Res.* **8**, 370–378 (1998).
96. Autier, P. et al. Melanoma and use of sunscreens: an EORTC case-control study in Germany, Belgium and France. The EORTC Melanoma Cooperative Group. *Int. J. Cancer* **61**, 749–755 (1995).
97. Dennis, L. K., Beane Freeman, L. E. & VanBeek, M. J. Sunscreen use and the risk for melanoma: a quantitative review. *Ann. Intern. Med.* **139**, 966–978 (2003).
98. Huncharek, M. & Kupelnick, B. Use of topical sunscreens and the risk of malignant melanoma: a meta-analysis of 9067 patients from 11 case-control studies. *Am. J. Public Health* **92**, 1173–1177 (2002).
99. Palmer, J. S. et al. Melanocortin-1 receptor polymorphisms and risk of melanoma: is the association explained solely by pigmentation phenotype? *Am. J. Hum. Genet.* **66**, 176–186 (2000).
100. Wenzel, E. et al. (Pheo)melanin photosensitizes UVA-induced DNA damage in cultured human melanocytes. *J. Invest. Dermatol.* **111**, 678–682 (1998).

Acknowledgements We thank S. R. Granter, M. E. Bigby, H. A. Haynes, A. B. Kimball, J. Rees, A. J. Sober, R. Stern and H. Tsao for useful comments and discussions. This work was supported by grants from the NIH and Doris Duke Charitable Foundation, and a Ruth L. Kirschstein National Research Service Award (J.Y.L.). D.E.F. is Distinguished Clinical Investigator of the Doris Duke Charitable Foundation and Jan and Charles Nirenberg Fellow in Pediatric Oncology at the Dana-Farber Cancer Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to D.E.F. (david_fisher@dfci.harvard.edu).

Melanoma biology and new targeted therapy

Vanessa Gray-Schopfer¹, Claudia Wellbrock¹ & Richard Marais¹

Melanoma is a cancer that arises from melanocytes, specialized pigmented cells that are found predominantly in the skin. The incidence of melanoma is rising steadily in western populations — the number of cases worldwide has doubled in the past 20 years. In its early stages malignant melanoma can be cured by surgical resection, but once it has progressed to the metastatic stage it is extremely difficult to treat and does not respond to current therapies. Recent discoveries in cell signalling have provided greater understanding of the biology that underlies melanoma, and these advances are being exploited to provide targeted drugs and new therapeutic approaches.

Melanocytes are specialized pigmented cells that are found predominantly in the skin and eyes, where they produce melanins, the pigments responsible for skin and hair colour. Cutaneous melanocytes originate from highly motile neural-crest progenitors that migrate to the skin during embryonic development. In the skin, melanocytes reside in the basal layer of the epidermis (Fig. 1) and in the hair follicles, and their homeostasis is regulated by epidermal keratinocytes¹. In response to ultraviolet (UV) radiation, keratinocytes secrete factors that regulate melanocyte survival, differentiation, proliferation and motility, stimulating melanocytes to produce melanin and resulting in the tanning response. Thereby, melanocytes have a key role in protecting our skin from the damaging effects of UV radiation and in preventing skin cancer (see page 843). Consequently, people lacking functional melanocytes in pigmentary disorders such as vitiligo and albinism are hypersensitive to UV radiation². It might seem perverse, therefore, that melanocytes are the precursors of melanoma, the most deadly form of skin cancer.

Skin cancer is the third most common human malignancy and its global incidence is rising at an alarming rate, with basal cell carcinoma, squamous cell carcinoma and melanoma being the most common forms. There are an estimated 2–3 million cases of skin cancer across the world each year, and although melanoma only accounts for about 132,000 of these (World Health Organization), it is the most dangerous form, accounting for most skin cancer deaths. If melanoma is diagnosed early it can be cured by surgical resection, and about 80% of cases are dealt with in this way. However, metastatic malignant melanoma is largely refractory to existing therapies and has a very poor prognosis, with a median survival rate of 6 months and 5-year survival rate of less than 5%³, so new treatment strategies are urgently needed. Recent studies have provided a much improved understanding of melanoma biology and we review these findings and the exciting opportunities they are providing for new therapeutic approaches to this disease.

Types and progression of cutaneous melanoma

Mutations in critical growth regulatory genes, the production of auto-crine growth factors and the loss of adhesion receptors all contribute to disrupted intracellular signalling in melanocytes, allowing them to escape their tight regulation by keratinocytes⁴. Consequently, melanocytes can proliferate and spread, leading to formation of a naevus or common mole (Fig. 1). Melanocyte proliferation can be restricted to the epidermis (junctional naevus), the dermis (dermal naevus) or overlapping components of both (compound naevus). Naevi are generally benign but can progress to the radial-growth-phase (RGP) melanoma (Fig. 1), an intra-epidermal lesion that can involve some local micro-invasion of the dermis. RGP cells can progress to the vertical-growth-

phase (VGP), a more dangerous stage in which the cells have metastatic potential, with nodules or nests of cells invading the dermis (Fig. 1). Not all melanomas pass through each of these individual phases — RGP or VGP can both develop directly from isolated melanocytes or naevi, and both can progress directly to metastatic malignant melanoma⁵.

There are four main clinical subtypes of melanoma⁶. Nodular melanoma consists of raised nodules without a significant flat portion. Acral lentiginous melanoma (ALM) tends to be found on the palms of the hands, the soles of the feet and in the nail bed, and so is not associated with UV exposure. ALM accounts for about 50% of melanoma in non-Caucasian populations⁷. Lentigo maligna is generally flat in appearance and occurs on sun-exposed regions in the elderly. It is therefore associated with lifetime chronic sun exposure. Finally, superficial spreading melanoma (SSM) is by far the most common form of melanoma. It is usually flat with an intra-epidermal component, particularly at the edges, and is linked to episodes of severe sunburn, especially at an early age. SSM is the third most common cancer in young people in the UK and USA, and also accounts for the remaining 50% of melanoma cases in non-Caucasians^{8,9}. So far, melanoma diagnosis has been based on pathology, but Bastian and colleagues recently demonstrated that genome-wide alterations in DNA copy number together with analysis of individual somatic mutation can be used to distinguish between the different melanoma subtypes with 70% accuracy¹⁰. These data provide new prognostic and diagnostic opportunities and suggest that the different melanomas develop along distinct genetic pathways.

Insight into melanoma biology from cell signalling pathways

Melanoma is a complex genetic disease, the management of which will require an in-depth understanding of the biology underlying its initiation and progression. This will allow improved staging and subtype classification, and will lead to the design of better therapeutic agents and approaches. Comprehensive strategies such as comparative genomic hybridization and mutation analysis by gene resequencing have identified some of the crucial cell-signalling pathways in this disease, which are discussed below.

Ras/Raf/MEK/ERK signalling

The Ras/Raf/MEK/ERK pathway regulates cell fate decisions downstream of receptor tyrosine kinases, cytokines and heterotrimeric G-protein-coupled receptors¹¹. The small G protein Ras is attached to the inner leaflet of the plasma membrane, whereas Raf, MEK and ERK are cytosolic protein kinases (Fig. 2). In melanocytes, this pathway is activated by growth factors such as stem-cell factor (SCF), fibroblast growth factor (FGF) and hepatocyte growth factor (HGF)¹², but

¹The Institute of Cancer Research, Signal Transduction Team, Cancer Research UK Centre of Cell and Molecular Biology, 237 Fulham Road, London SW3 6JB, UK.

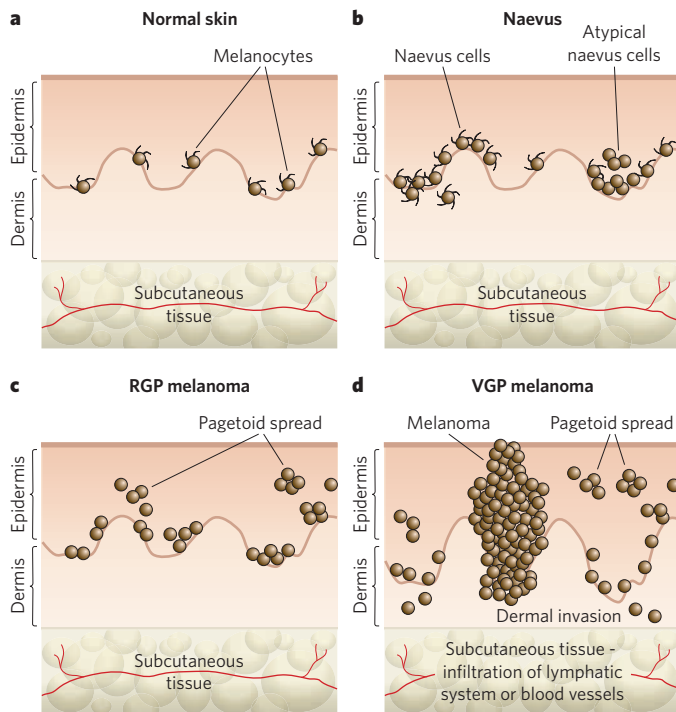


Figure 1 | Progression of melanocyte transformation. There are various stages of melanocytic lesion, each of which is marked by a new clone of cells with growth advantages over the surrounding tissues. **a**, Normal skin. This shows an even distribution of dendritic melanocytes within the basal layer of the epidermis. **b**, Naevus. In the early stages, benign melanocytic naevi occur with increased numbers of dendritic melanocytes. According to their localization, naevi are termed either junctional, dermal or compound. Some naevi are dysplastic, with morphologically atypical melanocytes. **c**, Radial-growth-phase (RGP) melanoma. This is considered to be the primary malignant stage. **d**, Vertical-growth-phase (VGP) melanoma. This is the first stage that is considered to have malignant potential and leads directly to metastatic malignant melanoma, the most deadly stage, by infiltration of the vascular and lymphatic systems. Pagetoid spread describes the upward migration or vertical stacking of melanocytes that is a histological characteristic of melanoma.

individually these growth factors induce only weak or transient ERK activation, leading to a modest mitogenic effect. Thus for proliferation the combined action of several growth factors is required to stimulate strong and sustained ERK activity in melanocytes^{12,13}.

This pathway is also a key regulator of melanoma cell proliferation, with ERK being hyperactivated in up to 90% of human melanomas¹⁴. In melanoma, ERK can be activated by the production of autocrine growth factors¹⁵ or in rare cases by mutational activation of growth-factor receptors such as c-Kit¹⁶. A more common mechanism is through gain-of-function mutations in *NRAS*, one of three *Ras* genes in humans (the others being *HRAS* and *KRAS*). *NRAS* is mutated in between 15% and 30% of melanomas, the most common substitution being leucine for glutamine at position 61 (Q61L)¹⁷ (Table 1). Importantly, oncogenic *Ras* can induce melanoma in p16^{INK4a}-deficient mice and it also has a role in tumour maintenance^{18,19}.

The most commonly mutated component of this pathway is *BRAF*, one of the three human *Raf* genes (together with *ARAF* and *CRAF*). *BRAF* is mutated in 50% to 70% of melanomas¹⁷ (Table 1), and the most common mutation is a glutamic acid for valine substitution at position 600 (V600E)¹⁷. V600E *BRAF* stimulates constitutive ERK signalling, stimulating proliferation and survival and providing essential tumour growth and maintenance functions²⁰. V600E *BRAF* also contributes to neo-angiogenesis by stimulating autocrine vascular endothelial growth factor (VEGF) secretion²¹ and recent studies have identified several genes in melanoma that function downstream of V600E *BRAF*. These include those

encoding the transcription factors MITF (microphthalmia-associated transcription factor)²² and BRN-2 (POU domain class 3 transcription factor)²³, the cell cycle regulators cyclin D1 (ref. 24) and p16^{INK4a} (refs 25, 26), and the tumour maintenance enzymes matrix metalloproteinase-1 (ref. 27) and inducible nitric oxide synthase²⁸. Thus, *BRAF* is implicated in several aspects of melanoma induction and progression.

Ras-ERK and phosphoinositide-3-OH kinase signalling

Another signalling pathway that is emerging as important in melanoma is the phosphoinositide-3-OH kinase (PI(3)K) pathway (Fig. 2). Phosphoinositides are membrane lipids that are converted to second messengers through hyper-phosphorylation by one of a family of PI(3)Ks²⁹. The lipid second messengers activate numerous downstream effector pathways, and signalling is terminated by the lipid phosphatase PTEN (phosphate and tensin homologue). PI(3)K signalling regulates cell survival, proliferation, growth (increase in cell mass) and motility, and it is hyperactivated in a high proportion of melanomas (Table 1). This is explained in part by the findings that PI(3)K mutations occur in 3% of metastatic melanomas³⁰, PTEN function is lost in between 5% and 20% of late-stage melanomas³¹, and the PI(3)K effector protein kinase B (PKB, also known as Akt) is overexpressed in up to 60% of melanomas³².

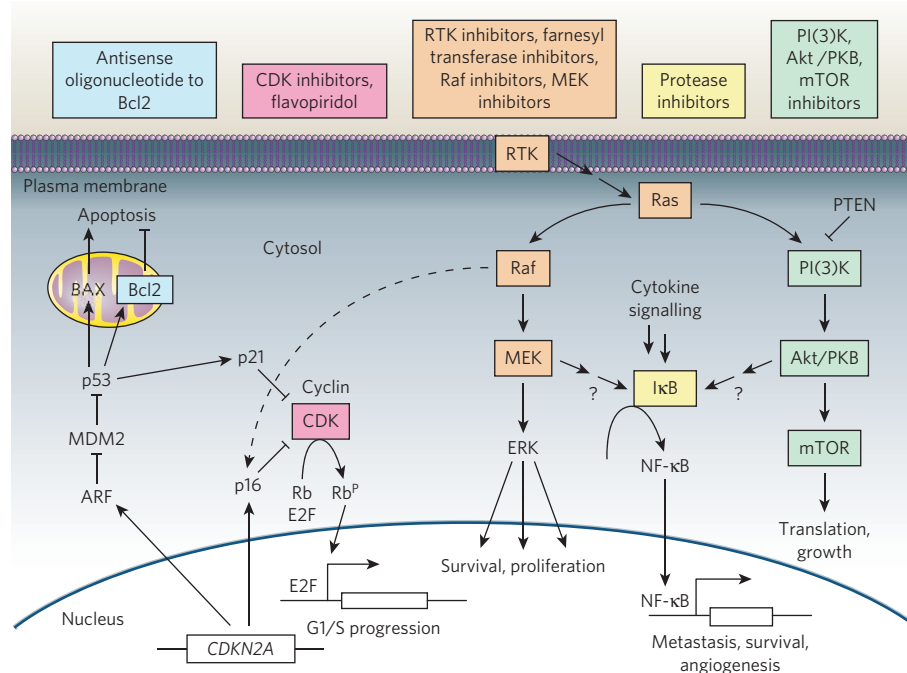
In three-dimensional melanoma cultures, ERK and PI(3)K signalling must both be inhibited to suppress cell growth, demonstrating that both pathways are important³³. Notably, in melanoma, *NRAS* and *BRAF* mutations are mutually exclusive, as are *NRAS* and *PTEN* mutations, whereas *BRAF* and *PTEN* mutations are coincident in about 20% of cases³⁴. The epistatic relationship of these genes has not been shown genetically, but these mutation patterns conform neatly to our current understanding of the pathway (Fig. 2). *BRAF* and PI(3)K are activated downstream of *Ras*, so in the presence of oncogenic *NRAS* additional mutations in *BRAF* and *PTEN* are unnecessary, because these components are already activated. Indeed, their mutation in this context might even antagonize tumour progression, because they could stimulate pathway superactivation. By contrast, because *PTEN* and *BRAF* are not mutually activating and these pathways cooperate to stimulate proliferation, separate mutations are required to activate both pathways. Support for this model comes from the observation that oncogenic *Ras* transforms melanocytes more efficiently than oncogenic *BRAF*³⁵, presumably because of the PI(3)K component. It is curious, therefore, that *BRAF* mutations are more prevalent than *Ras* mutations in melanoma, and this presumably reflects other genetic or biological pressures that favour the progression of melanomas with *BRAF* mutations over those with mutated *NRAS*.

Microphthalmia-associated transcription factor

MITF is a basic helix–loop–helix leucine zipper transcription factor that is considered to be the master regulator of melanocyte biology, partly because it regulates the expression of melanogenic proteins such as tyrosinase, silver homologue (GP100) and melanoma-associated antigen recognized by T cells-1 (MART-1, also known as melan-A)³⁶ (see also page 843). However, MITF also regulates melanoblast survival and melanocyte lineage commitment, and is a key player in melanoma³⁶. MITF is expressed in most human melanomas and its target genes are diagnostic markers for this disease³⁶. Furthermore, continued MITF expression is essential for melanoma cell proliferation and survival³⁶. However, we propose that the protein levels must be carefully controlled. MITF is expressed at significantly lower levels in melanoma cells than in melanocytes, and increased MITF levels reduce melanoma cell proliferation even in the presence of oncogenic *BRAF*³⁷. We posit that MITF regulates distinct functions in melanocytic cells at different levels of expression. High levels predispose to cell cycle arrest and differentiation, presumably by controlling cell cycle regulators (such as p16^{INK4a}, cyclin-dependent kinase 2 (Cdk2) and p21^{Cip1})³⁶, whereas critically low levels lead to cell cycle arrest and apoptosis. Only at intermediate levels is proliferation favoured (Fig. 3).

Importantly, recent studies have shown that MITF can cooperate with V600E *BRAF* to transform melanocytes immortalized by expression of the

Figure 2 | Melanoma drugs and their target pathways. Drugs developed to treat melanoma (top) and their targets (below) in several of the main pathways involved in melanoma initiation, progression and maintenance are indicated by different colours. One such pathway is the three-tiered Ras/Raf/MEK mitogen-activated protein kinase (MAPK) signalling module. Another is the PI(3)K pathway, which regulates cell survival, proliferation, growth and motility, and is inhibited by the lipid phosphatase PTEN. The involvement of signalling by NF- κ B, which transcribes various genes including matrix metalloproteases (which are involved in metastasis), IAPs and Bcl-2 (anti-apoptotic factors), and VEGF and cytokines (which partake in angiogenesis), is also shown. However, upstream NF- κ B signals are omitted for simplicity. The schematic includes interactions between MAPK, PI(3)K and NF- κ B signalling. The p16^{INK4a}/RB and ARF signalling pathways are also important, as their disruption can lead to melanoma by preventing melanocyte senescence.



catalytic subunit of telomerase (TERT), dominant-negative p53 and activated Cdk4 (ref. 38). These data show that although MITF alone cannot transform normal or immortalized human melanocytes, it can cooperate with ^{V600E}BRAF to contribute to the transformation process, functioning as a lineage-specific 'addictive oncogene'³⁸, presumably because it provides essential survival functions and contributes to proliferation. In this context, and bearing in mind that ERK is hyperactivated in melanoma and required for proliferation and survival, it is striking that MITF is targeted for degradation after its phosphorylation by ERK³⁹. Indeed, constitutive activation of ERK by ^{V600E}BRAF in melanocytes results in constant downregulation of MITF^{22,38,39} (Fig. 3). Therefore, to counteract a situation in which MITF levels can become critically low, melanoma cells seem to develop strategies to recover MITF protein to levels that are compatible with tumour progression.

It has recently been shown that *MITF* is amplified in a small percentage (10–16%) of metastatic melanomas in which *BRAF* is mutated³⁸ (Table 1). Surprisingly, despite the fact that the gene can be amplified 10–100-fold, MITF levels are only increased about 1.5-fold compared with cells without amplification³⁸, again suggesting that MITF levels must be maintained within narrow limits. However, because only 10–16% of *BRAF*-mutated melanomas have *MITF* amplification, this raises the crucial question of how the remaining 84–90% counteract MITF degradation mediated by hyperactivated ERK. One mechanism could involve β -catenin, which can induce *MITF* expression through a LEF-1/TCF binding site in the *MITF* promoter⁴⁰ and, although stabilizing mutations in β -catenin are rare in melanoma⁴¹, nuclear and/or cytoplasmic localization of β -catenin was found in 28% of metastatic melanoma⁴².

Other signalling pathways

Finally, it is worth noting that several genetic alterations common in melanoma seem to reduce apoptosis, including the mutations in *NRAS* and *BRAF*, the overexpression of B-cell leukaemia/lymphoma 2 (*Bcl-2*), nuclear factor- κ B (NF- κ B) and *Akt3*, and loss of *PTEN* (Fig. 2). Unsurprisingly, some of these pathways also promote growth, because proliferation and survival pathways overlap. Other survival signals may come from β -catenin pathway upregulation through loss or silencing of adenomatous polyposis coli (*APC*)⁴³, or from rare activating mutations in β -catenin (*CTNNB1*)⁴¹. Overexpression of the transcription factor SKI might also have a role⁴⁴, as might the alterations that affect p14^{ARF}. Finally, the frequent silencing of apoptotic peptidase activating factor-1 (APAF-1) in advanced melanomas⁴⁵ might also contribute to

survival. Clearly these pathways warrant further study to determine their role in human melanoma and the potential they offer for therapeutic intervention.

Overriding senescence mechanisms

Normal somatic cells have restricted proliferative potential, and after a finite number of divisions they exit from the cell cycle and enter a state known as senescence⁴⁶. Senescence is a key cellular protection mechanism against cancer because it halts aberrant cell proliferation, and the process is physiologically relevant because it is induced by activated oncogenes in normal somatic cells. Cancer cells override senescence by inactivating key pathways such as those regulating p16^{INK4a}/Rb (retinoblastoma protein) and p53 (ref. 47).

The tumour suppressor protein p16^{INK4a} functions by binding to the cyclin-D-dependent protein kinases Cdk4 and 6, blocking their activity and preventing Rb phosphorylation⁴⁸ (Fig. 2). The genes encoding p16^{INK4a} and Cdk4 are both high-penetrance loci for melanoma susceptibility. p16^{INK4a} is encoded by *CdkN2A*, a locus that also encodes the tumour suppressor p14^{ARF} (p19^{ARF} in mice) by means of a different promoter and through use of alternative reading frames⁴⁹. In familial melanoma, *CdkN2A* mutations generally affect only p16^{INK4a} or p16^{INK4a} together with p14^{ARF}, only in rare cases is p14^{ARF} alone affected. Mutations in *Cdk4* disrupt p16^{INK4a} binding⁴⁹. Somatic mutations in these genes are also found in most sporadic melanomas. p16^{INK4a} is inactivated by point mutations, deletions, promoter methylation⁵⁰ (Table 1) or through

Table 1 | Selected genetic alterations in malignant melanomas

Gene type	Gene	Alteration frequency/type(s) in melanoma (%)	References
Oncogenes	<i>BRAF</i>	50–70% mutated	17
	<i>NRAS</i>	15–30% mutated	17, 34
	<i>AKT3</i>	Overexpressed	32
Tumour suppressors	<i>CDKN2A</i>	30–70% deleted, mutated or silenced	50
	<i>PTEN</i>	5–20% deleted or mutated	31, 34
	<i>APAF-1</i>	40% silenced	45
	<i>p53</i>	10% lost or mutated	89
Others	Cyclin D1	6–44% amplified	52
	<i>MITF</i>	10–16% amplified	38

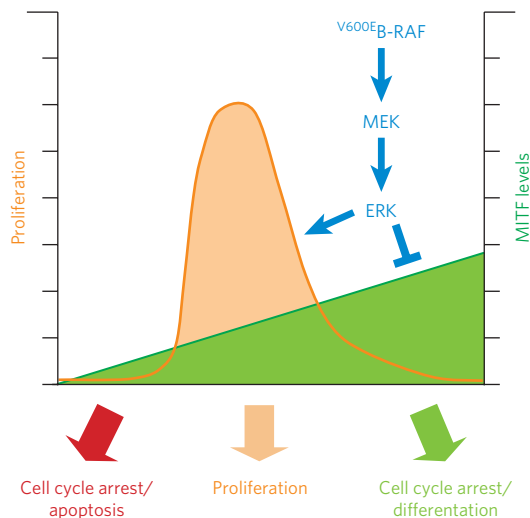


Figure 3 | Model of MITF regulation in melanoma. Increased MITF expression levels inhibit proliferation and favour differentiation. Suppression of MITF levels allows optimal proliferation. In melanoma cells this is achieved by constitutive activation of ERK through oncogenic BRAF. Active ERK is also required for melanoma cell proliferation. Complete loss of MITF expression results in cell cycle arrest and/or apoptosis. The general mechanisms regulating MITF levels to ensure they are compatible with maximum cell proliferation are unknown.

transcriptional silencing by overexpression of the transcriptional suppressor, inhibitor of differentiation 1 (ID1)⁵¹. Alternatively, *Cdk4* mutations that block p16^{INK4a} binding can occur, or cyclin D1 can be overexpressed (Table 1), in part as a result of gene amplification⁵². Furthermore, most cultured melanoma cell lines seem to have a defect in p16^{INK4a} or Rb, or overexpression of *Cdk4* or cyclin D1. Clearly, p16^{INK4a}/Rb signalling is a key regulator of melanocyte senescence and, consequently, to override senescence melanoma cells must inactivate this pathway⁴⁶.

One intriguing observation is that *BRAF* is mutated in up to 80% of benign naevi⁵³, yet most naevi remain indolent for decades and only rarely progress to melanoma. Thus, oncogenic BRAF alone is insufficient for melanoma initiation and, strikingly, V600E-BRAF was recently found to induce p16^{INK4a} expression and senescence in primary human melanocytes *in vitro*^{25,26} (Fig. 2). Notably, V600E-BRAF cannot transform human melanocytes, even when they are immortalized by TERT expression³⁸, but it does transform p16^{INK4a}-deficient mouse melanocytes³⁵, and p16^{INK4a} deficiency is required for Ras-induced melanoma in mice¹⁹. Furthermore, p16^{INK4a} upregulation seems to contribute to the induction of senescence in naevi^{25,26} and p16^{INK4a} defects are found in dysplastic but not benign naevi⁵⁴.

Together, these studies suggest an intriguing model whereby p16^{INK4a} serves as a brake to V600E-BRAF-triggered melanocyte transformation. In this model, acquisition of mutations in *BRAF* or *NRAS* would lead to hyperproliferation and consequent p16^{INK4a}-mediated senescence, generating naevi that contain clones of senescent melanocytes. These studies suggest that melanoma progression requires p16^{INK4a}/Rb pathway inactivation. Curiously, however, p16^{INK4a} is expressed in a mosaic pattern in naevi, whereas the expression of β -galactosidase, another marker of senescence, is uniform^{25,26}. It is possible that p16^{INK4a} arrests proliferation in some cells at levels that are below the sensitivity of immunostaining approaches, explaining why expression is not uniform. Another possibility is that senescent cells secrete senescence-inducing agents that affect neighbouring cells that do not express p16^{INK4a}. Finally, it is feasible that widespread chromatin remodelling and gene silencing due to Rb-family protein hyperactivation⁵⁵ causes the cells to become so transcriptionally inactive that p16^{INK4a} becomes silenced in some of the cells, but growth stasis is maintained by the silencing of genes that are required for proliferation.

In addition, senescence might be induced by other factors. Consistent with this idea is the fact that patients with homozygous p16^{INK4a} deletions still develop benign naevi. Indeed, humans carrying p16^{INK4a} mutations not only have increased susceptibility to melanoma but usually also have many moles, which are often large⁴⁶, implying that p16^{INK4a} is not the only inducer of senescence in naevi. Other mechanisms might involve p53, as seen in cultured melanocytes from p16^{INK4a}-deficient patients⁴⁶. Notably, pigmented melanophages (macrophages that have ingested melanin) and lymphocytes are often found near dermal p53-positive cells, which is indicative of melanocyte death. This suggests that a p53-dependent 'back-up' (also known as 'crisis') senescence checkpoint can be activated in melanocytes after p16^{INK4a}-dependent senescence (the principal mechanism protecting melanocytes from NRAS- or BRAF-mediated transformation) has been evaded²⁵. Of course, it is possible that other pathways that do not involve p16^{INK4a}/Rb- or p53-dependent mechanisms also have a role.

Treating melanoma

Melanoma is an extremely aggressive disease with high metastatic potential and a notoriously high resistance to cytotoxic agents. This is thought to be because melanocytes originate from highly motile cells that have enhanced survival properties. Melanoma cells have low levels of spontaneous apoptosis *in vivo* compared with other tumour cell types, and are relatively resistant to drug-induced apoptosis *in vitro*⁵⁶. Most chemotherapeutic drugs function by inducing apoptosis in malignant cells, so resistance to apoptosis is likely to underlie drug resistance in melanoma⁵⁶, and this extraordinary resistance to chemotherapy, radiotherapy and immunotherapy is a major barrier to successful treatment of melanoma.

There are several approved postoperative adjuvant therapies for malignant melanoma⁵⁷. Interferon- α (IFN- α) is the most commonly used adjuvant immunotherapy for advanced melanoma, although its efficacy is still a matter of debate. High-dose interleukin-2 (IL-2) has also been approved, but response rates are low and toxicity is a problem. Dacarbazine (DTIC) is the reference approved chemotherapeutic agent for the treatment of advanced melanoma, and drugs such as carmustine (BiCNU), paclitaxel (Taxol), temozolomide and cisplatin have shown single-agent activity in metastatic disease⁵⁷. In addition, many different immunotherapies have been tested, but so far none of these approaches has reached regulatory approval⁵⁸. However, recent trials of genetically engineered lymphocyte transfer have shown promise, with 2 (out of 17) patients achieving full remission for 18 months⁵⁹. Unfortunately, these therapies all contribute little to overall patient survival, so the identification of the signalling pathways that are central to melanoma initiation and progression is opening an exciting new area for melanoma treatment. It provides the opportunity to develop targeted therapies and to tailor those treatments to patients according to the genetic lesions that underlie their individual disease. On the basis of these principles, several new targeted agents are currently being developed and tested alone and in combination with conventional chemotherapies.

Targeted therapies

The most attractive cancer targets are those on which the cancer cells have come to depend for progression. This concept of 'oncogene addiction' argues that cancer cells rely more heavily on hyperactivated pathways than do normal cells, and therefore rely on the activated oncogenes that drive those pathways. This provides a therapeutic opportunity, because cancer cells would be more sensitive to inhibition of those pathways/oncogenes than normal cells. Not all oncogenes are tractable targets, but enzymes such as kinases, proteases and phosphatases are highly valued because their catalytic sites include deep pockets into which carefully designed drugs can bind with a reasonable margin of selectivity. Because NRAS and BRAF are validated therapeutic targets in melanoma, it is not surprising that drugs that target this pathway are of considerable interest. The first drugs targeting this pathway to enter the clinic were the Ras farnesyl transferase inhibitors (Table 2), which were designed to block an essential post-translational modification on Ras.

Although these drugs are not very specific, they may have some value in combination with cisplatin⁶⁰, although the responses are unlikely to be mechanism-based.

The multi-kinase inhibitor sorafenib (BAY 43-9006) targets BRAF, CRAF and the VEGF and platelet-derived growth factor (PDGF) receptor tyrosine kinases⁶¹ (Table 2). As a monotherapy, sorafenib shows only modest activity against melanoma, but when it is combined with carboplatin and paclitaxel the results are more encouraging⁶². It is not clear why sorafenib monotherapy is so inefficient in melanoma. It may be that it is not potent enough, or it is possible that melanoma cell proliferation is driven by alternative pathways when Raf/MEK/ERK signalling is blocked⁶³. It is also possible that other survival mechanisms mediated by tumour-necrosis factor- α (TNF- α) provide a survival 'escape' route, allowing melanoma cells to survive when BRAF signalling is inhibited⁶⁴ or that feedback mechanism(s) could negate the effect of sorafenib⁶⁵. Clearly, more potent BRAF drugs are required to test whether BRAF is a suitable target in melanoma. To this end, the BRAF inhibitor CHIR-265 recently entered clinical trials for melanoma (Table 2).

Although BRAF is the current focus of Raf drug development, the other Raf isoforms should not be overlooked, not least because BRAF activates CRAF in mammalian cells⁶⁶ and, although CRAF is not required for MEK activation in the presence of ^{V600E}BRAF, it is required for proliferation²⁰. Furthermore, we recently discovered that when Ras is mutated in melanoma, it is CRAF and not BRAF that activates MEK⁶⁷. So CRAF is involved in melanoma cell proliferation, and although it is currently unclear whether these functions require CRAF kinase activity, the data do suggest that panspecific Raf drugs would be better anti-melanoma agents than BRAF-specific drugs.

As an alternative, drugs such as PD0325901 and AZD6244, which target MEK, have been developed (Table 2). These inhibit *in vitro* proliferation, soft-agar colony formation and matrigel invasion of ^{V600E}BRAF mutant human melanoma cells, and are also effective against ^{V600E}BRAF melanoma xenografts⁶⁸. Notably, melanoma cells in which BRAF is mutated are more sensitive to MEK inhibitors than cells in which Ras is mutated⁶⁹. This is unexpected, because Ras and BRAF both signal through MEK. It suggests either that activation of additional pathways by Ras makes melanoma cells less dependent on MEK or, conversely, that BRAF-mutant cells are 'more addicted' to MEK signalling, implying that BRAF-mutant tumours will respond more readily to MEK inhibition than Ras-mutant tumours. AZD6244 and PD0325901 are currently being

tested in clinical trials in patients with advanced melanoma (Table 2). Finally, the anthrax lethal toxin, which selectively degrades and inactivates MEK1 and MEK2 (ref. 70), is also being tested in melanoma clinical trials. However, it should be noted that one potential problem of MEK inhibition is that MEK is also activated by tumour progression locus 2 (TPL2), a protein kinase that regulates innate immunity⁷¹, so chronic MEK inhibition might have negative effects on innate immunity. This would not be a problem with Raf inhibition, and is a factor that should be taken into consideration during clinical trial design.

PI(3)K signalling is also likely to be a good target for the treatment of melanoma, and agents that target PI(3)K, PKB and other downstream components of this pathway such as mammalian target of rapamycin (mTOR) are being developed. Among these, mTOR inhibitors such as CCI-779 or RAD001 are the most advanced⁷² (Table 2). The importance of this pathway to melanoma suggests that drugs targeted to this pathway could be beneficial in correctly selected patients, particularly if used in combination with other agents. In particular, because Ras-ERK and PI(3)K/PKB signalling both contribute to melanoma cell survival and could in part explain the notorious resistance of melanoma cells to cytotoxic drugs, it would be particularly interesting to test anti-Raf-MEK drugs together with anti-PI(3)K/PKB drugs, with or without agents that induce apoptosis. Other drugs that could be considered are those targeting the survival signals provided by NF- κ B signalling (sensitive to velcade and aspirin) or survival factors such as Bcl-2. Bcl-2 in particular is an important melanocyte and melanoma cell survival factor (Fig. 2) that might protect cells from toxic melanin intermediates or the reactive oxygen species generated during melanogenesis and/or UV irradiation. Normal melanocytes and melanoblasts express high levels of Bcl-2 (ref. 73) and melanocyte numbers decrease prematurely in *Bcl2*-null mice owing to loss of the stem-cell population⁷⁴. In preclinical studies, antisense oligonucleotides against Bcl-2 (such as oblimersen) can sensitize melanoma cells to chemotherapy⁷⁵ (Table 2) and a Phase III study has shown some benefit in progression-free survival when used in combination with dacarbazine^{57,76}. A bispecific oligonucleotide to Bcl-2 and Bcl-X_L is also being tested⁷⁷.

The fact that p16^{INK4a} function is lost in most melanomas identifies Cdk4 and 6 as other potential therapeutic targets. Pan CDK inhibitors such as flavopiridol are largely inactive and there is no evidence of clinical activity in malignant melanoma, but these agents could be effective in some patients, because CDK inhibition preferentially suppresses

Table 2 | Selected drugs for targeted melanoma therapy

Drug	Function	Target(s)	Study phase	References
BAY 43-9006 (sorafenib)	Kinase inhibitor	VEGFR, PDGFR, Raf kinases	II	63
CHIR-265	Kinase inhibitor	Raf kinases, VEGFR	I	-
PD0325901	Kinase inhibitor	MEK	I/II	-
AZD6244	Kinase inhibitor	MEK	II	90-92
Oblimersen	Antisense oligonucleotide	Bcl-2	III (with dacarbazine)	76
SU11248 (sunitinib)	Kinase inhibitor	VEGFR, PDGFR, Kit, FLT3	III (renal cancer)	93
SU5416 (semaxanib)	Kinase inhibitor	VEGFR	II	94
AG013736	Kinase inhibitor	VEGFR, PDGFR, Kit	II (renal cancer)	-
PD0332991	Kinase inhibitor	Cdk4/6	I (neoplasms, non-Hodgkin's lymphoma)	-
CCI-779 (temsirolimus)	mTOR inhibitor	mTOR, PTEN/PI(3)K/Akt pathway	II	95
RAD001 (everolimus)	mTOR inhibitor	mTOR, PTEN/PI(3)K/Akt pathway	I	96
R115777 (tibifarnib)	Farnesyl transferase inhibitor	Ras	Preclinical	97
PS-341 (bortezomib, velcade)	Proteasome inhibitor	NF- κ B	II	98
BMS-345541	Kinase inhibitor	I κ B	Preclinical	99
Vitaxin	Humanized antibody to $\alpha_v\beta_3$	$\alpha_v\beta_3$ integrin	II	85
EMD 121974 (cilengitide)	Angiogenesis inhibitor	Integrin	II	85
17AAG	Hsp90 inhibitor	Hsp90	II	79
ZSTK474	PI(3)K inhibitor	PI(3)K	Preclinical	100

FLT3, FMS-like tyrosine kinase 3; PDGFR, PDGF receptor; VEGFR, VEGF receptor.

expression of mRNAs with short half-lives, including those encoding antiapoptotic proteins such as X-linked inhibitor of apoptosis (XIAP). More selective Cdk4/6 drugs, such as PD0332991 (Table 2), which is active in colon carcinoma xenograft models⁷⁸, might be more effective. The inherent redundancy in this pathway does not seem to be an issue, because Cdk4/6 inhibition induces G1 cell cycle arrest and tumour regression *in vivo*, whereas Cdk1/2 inhibition induces E2F-dependent cell death, demonstrating that the activities of these two related kinases can be separated. Once again, these drugs could be combined with other agents to provide opportunities for melanoma treatment, and it would be particularly interesting to combine them with Ras-ERK pathway-targeting drugs.

Targeting multiple pathways

Another approach is to use agents that target several pathways together. The molecular chaperone heat-shock protein 90 (Hsp90) regulates folding and function of newly synthesized proteins, among which are many protein kinases, including CRAF, BRAF, Cdk4 and Cdk6 (ref. 78). Hsp90 inhibition by the benzoquinone ansomycin 17AAG (Table 2) causes degradation of its client proteins through the ubiquitin-dependent proteasomal pathway. Thus, Hsp90 inhibition can target several pathways simultaneously and, surprisingly, this does not lead to significant levels of toxicity as might be expected⁷⁹. Importantly, ^{V600E}BRAF is more sensitive to 17AAG than wild-type BRAF, but 17AAG also targets wild-type BRAF and CRAF downstream of oncogenic Ras^{80,81}. So, however it is activated, 17AAG targets this pathway and, furthermore, because 17AAG targets many pathways, the presence of mutant BRAF does not increase melanoma cell sensitivity to 17AAG^{80,81}. Instead, 17AAG seems to obtain its therapeutic effect by targeting several pathways that are hyperactivated in cancer, but not normal, cells. Notably, responses to 17AAG have been seen in melanoma patients⁷⁹ and further trials are ongoing (www.clinicaltrials.gov).

Similarly, the proteasome inhibitor PS-341 (bortezomib, velcade; Table 2) has been shown to enhance activity of the drug temozolomide against melanoma *in vitro* and *in vivo*⁸², leading to clinical trials. The anti-tumour activity of PS-341 may be due to NF- κ B inhibition, but NF- κ B has not yet been fully validated as a target in melanoma. However, the availability of drugs such as BMS-345541, which is a selective I κ B kinase inhibitor⁸³, now make it possible to test the importance of NF- κ B as a target in melanoma.

An additional potential target is angiogenesis, which has a critical role in melanoma. Angiogenic factors such as VEGF, IL-8, basic FGF and PDGF are released by melanoma and host cells into the tumour microenvironment⁸⁴ and the VEGF receptor inhibitors SU5416 and AG013736 (Table 2) have anti-tumour activity in mice bearing human melanoma xenografts. However, there are no encouraging Phase II data, and because sorafenib, which is thought to work through an anti-angiogenic mechanism, provides only a marginal response in melanoma, it is possible that targeting angiogenesis alone will be an ineffective treatment for melanoma. In line with this is the observation that the VEGF receptor inhibitor EMD 121974 (Table 2) showed promise in preclinical studies but disappointing results in patients⁸⁵. It will be important to determine whether the sorafenib contribution in the ongoing combination studies⁶² is due to its anti-angiogenic activity as this will indicate whether anti-angiogenic drugs are worthy of a role in combination therapy approaches.

Finally, several other potential anti-melanoma drugs are in the pipeline, or are already available, but for some of these the targets have not been fully validated. Proteins such as the integrin $\alpha_3\beta_3$, which has an important role in melanoma cell adhesion, can be targeted by Vitaxin (Table 2), a humanized monoclonal antibody⁸⁶. The receptor tyrosine kinase c-Kit, which is activated by mutations in about 2% of metastatic melanomas, is inhibited by imatinib¹⁶. The c-Met and the FGF receptors should also not be overlooked, and although important factors such as MITF might not be suitable drug targets themselves, the processes that they control could involve downstream components that would be excellent targets. Our improved understanding of metastasis⁸⁷, cancer stem cells and our ability to use comprehensive analytical approaches to

identify new targets such as NEDD9 (neural-precursor-cell-expressed downregulated 9)⁸⁸ are also providing a range of new targets. As we learn more about the underlying biology of melanoma and our ability to screen and select patients increases, the opportunities for targeted clinical intervention will become extensive. Indeed, our improved understanding of melanoma biology is already defining new and exciting therapeutic approaches.

1. Slominski, A., Tobin, D. J., Shibahara, S. & Wortsman, J. Melanin pigmentation in mammalian skin and its hormonal regulation. *Physiol. Rev.* **84**, 1155–1228 (2004).
2. Boissy, R. E. & Nordlund, J. J. Molecular basis of congenital hypopigmentary disorders in humans: a review. *Pigment Cell Res.* **10**, 12–24 (1997).
3. Cummins, D. L. *et al.* Cutaneous malignant melanoma. *Mayo Clin. Proc.* **81**, 500–507 (2006).
4. Haass, N. K., Smalley, K. S. & Herlyn, M. The role of altered cell-cell communication in melanoma progression. *J. Mol. Histol.* **35**, 309–318 (2004).
5. Miller, A. J. & Mihm, M. C. Melanoma. *N. Engl. J. Med.* **355**, 51–65 (2006).
6. Clark, W. H. *et al.* A study of tumor progression: the precursor lesions of superficial spreading and nodular melanoma. *Hum. Pathol.* **15**, 1147–1165 (1984).
7. Kuchelmeister, C., Schaumburg-Lever, G. & Garbe, C. Acral cutaneous melanoma in caucasians: clinical features, histopathology and prognosis in 112 patients. *Br. J. Dermatol.* **143**, 275–280 (2000).
8. Ishihara, K., Saida, T. & Yamamoto, A. Updated statistical data for malignant melanoma in Japan. *Int. J. Clin. Oncol.* **6**, 109–116 (2001).
9. Gilchrist, B. A., Eller, M. S., Geller, A. C. & Yaar, M. The pathogenesis of melanoma induced by ultraviolet radiation. *N. Engl. J. Med.* **340**, 1341–1348 (1999).
10. Curtin, J. A. *et al.* Distinct sets of genetic alterations in melanoma. *N. Engl. J. Med.* **353**, 2135–2147 (2005).
11. Wellbrock, C., Karasides, M. & Marais, R. The RAF proteins take centre stage. *Nature Rev. Mol. Cell Biol.* **5**, 875–885 (2004).
12. Bohm, M. *et al.* Identification of p90^{RSK} as the probable CREB-Ser¹³³ kinase in human melanocytes. *Cell Growth Differ.* **6**, 291–302 (1995).
13. Wellbrock, C., Weisser, C., Geissinger, E., Troppmair, J. & Schartl, M. Activation of p59^{Fyn} leads to melanocyte dedifferentiation by influencing MKP-1-regulated mitogen-activated protein kinase signaling. *J. Biol. Chem.* **277**, 6443–6454 (2002).
14. Cohen, C. *et al.* Mitogen-activated protein kinase activation is an early event in melanoma progression. *Clin. Cancer Res.* **8**, 3728–3733 (2002).
15. Satyamoorthy, K. *et al.* Constitutive mitogen-activated protein kinase activation in melanoma is mediated by both BRAF mutations and autocrine growth factor stimulation. *Cancer Res.* **63**, 756–759 (2003).
16. Willmore-Payne, C., Holden, J. A., Tripp, S. & Layfield, L. J. Human malignant melanoma: detection of BRAF- and c-kit-activating mutations by high-resolution amplicon melting analysis. *Hum. Pathol.* **36**, 486–493 (2005).
17. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
18. Ackermann, J. *et al.* Metastasizing melanoma formation caused by expression of activated N-Ras^{Q61K} on an INK4a-deficient background. *Cancer Res.* **65**, 4005–4011 (2005).
19. Chin, L. *et al.* Essential role for oncogenic Ras in tumour maintenance. *Nature* **400**, 468–472 (1999).
20. Gray-Schopfer, V. C., da Rocha Dias, S. & Marais, R. The role of B-Raf in melanoma. *Cancer Metastasis Rev.* **24**, 165–183 (2005).
21. Sharma, A. *et al.* Mutant V599EB-Raf regulates growth and vascular development of malignant melanoma tumors. *Cancer Res.* **65**, 2412–2421 (2005).
22. Wellbrock, C. & Marais, R. Elevated expression of MITF counteracts B-Raf-stimulated melanocyte and melanoma cell proliferation. *J. Cell Biol.* **170**, 703–708 (2005).
23. Goodall, J. *et al.* The Brn-2 transcription factor links activated BRAF to melanoma proliferation. *Mol. Cell Biol.* **24**, 2923–2931 (2004).
24. Bhatt, K. V. *et al.* Adhesion control of cyclin D1 and p27^{Kip1} levels is deregulated in melanoma cells through BRAF-MEK-ERK signaling. *Oncogene* **24**, 3459–3471 (2005).
25. Gray-Schopfer, V. C. *et al.* Cellular senescence in naevi and immortalisation in melanoma: a role for p16? *Br. J. Cancer* **95**, 496–505 (2006).
26. Michaloglou, C. *et al.* BRAF^{E600}-associated senescence-like cell cycle arrest of human naevi. *Nature* **436**, 720–724 (2005).
27. Huntington, J. T. *et al.* Overexpression of collagenase 1 (MMP-1) is mediated by the ERK pathway in invasive melanoma cells: role of BRAF mutation and fibroblast growth factor signaling. *J. Biol. Chem.* **279**, 33168–33176 (2004).
28. Ellerhorst, J. A. *et al.* Regulation of iNOS by the p44/42 mitogen-activated protein kinase pathway in human melanoma. *Oncogene* **25**, 3956–3962 (2006).
29. Shaw, R. J. & Cantley, L. C. Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature* **441**, 424–430 (2006).
30. Ornholdt, K., Krockel, D., Ringborg, U. & Hansson, J. Mutations of PIK3CA are rare in cutaneous melanoma. *Melanoma Res.* **16**, 197–200 (2006).
31. Wu, H., Goel, V. & Haluska, F. G. PTEN signaling pathways in melanoma. *Oncogene* **22**, 3113–3122 (2003).
32. Stahl, J. M. *et al.* Deregulated Akt3 activity promotes development of malignant melanoma. *Cancer Res.* **64**, 7002–7010 (2004).
33. Smalley, K. S. *et al.* Multiple signaling pathways must be targeted to overcome drug resistance in cell lines derived from melanoma metastases. *Mol. Cancer Ther.* **5**, 1136–1144 (2006).
34. Tsao, H., Goel, V., Wu, H., Yang, G. & Haluska, F. G. Genetic interaction between NRAS and BRAF mutations and PTEN/MMAC1 inactivation in melanoma. *J. Invest. Dermatol.* **122**, 337–341 (2004).
35. Wellbrock, C. *et al.* ^{V599E}B-Raf is an oncogene in melanocytes. *Cancer Res.* **64**, 2338–2342 (2004).
36. Levy, C., Khaled, M. & Fisher, D. E. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol. Med.* **12**, 406–414 (2006).

37. Selzer, E. *et al.* The melanocyte-specific isoform of the microphthalmia transcription factor affects the phenotype of human melanoma. *Cancer Res.* **62**, 2098–2103 (2002).
38. Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
39. Wu, M. *et al.* c-Kit triggers dual phosphorylations, which couple activation and degradation of the essential melanocyte factor Mi. *Genes Dev.* **14**, 301–312 (2000).
40. Takeda, K. *et al.* Induction of melanocyte-specific microphthalmia-associated transcription factor by Wnt-3a. *J. Biol. Chem.* **275**, 14013–14016 (2000).
41. Omholt, K., Platz, A., Ringborg, U. & Hansson, J. Cytoplasmic and nuclear accumulation of β -catenin is rarely caused by CTNNB1 exon 3 mutations in cutaneous malignant melanoma. *Int. J. Cancer* **92**, 839–842 (2001).
42. Rimm, D. L., Caca, K., Hu, G., Harrison, F. B. & Fearon, E. R. Frequent nuclear/cytoplasmic localization of β -catenin without exon 3 mutations in malignant melanoma. *Am. J. Pathol.* **154**, 325–329 (1999).
43. Worm, J., Christensen, C., Gronbaek, K., Tulchinsky, E. & Guldberg, P. Genetic and epigenetic alterations of the APC gene in malignant melanoma. *Oncogene* **23**, 5215–5226 (2004).
44. Chen, D. *et al.* SKI activates Wnt/ β -catenin signaling in human melanoma. *Cancer Res.* **63**, 6626–6634 (2003).
45. Soengas, M. S. *et al.* Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature* **409**, 207–211 (2001).
46. Bennett, D. C. Human melanocyte senescence and melanoma susceptibility genes. *Oncogene* **22**, 3063–3069 (2003).
47. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
48. Serrano, M., Hannon, G. J. & Beach, D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366**, 704–707 (1993).
49. Hayward, N. K. Genetics of melanoma predisposition. *Oncogene* **22**, 3053–3062 (2003).
50. Sharpless, E. & Chin, L. The INK4a/ARF locus and melanoma. *Oncogene* **22**, 3092–3098 (2003).
51. Polsky, D., Young, A. Z., Busam, K. J. & Alani, R. M. The transcriptional repressor of p16/INK4a, Id1, is up-regulated in early melanomas. *Cancer Res.* **61**, 6008–6011 (2001).
52. Sauter, E. R. *et al.* Cyclin D1 is a candidate oncogene in cutaneous melanoma. *Cancer Res.* **62**, 3200–3206 (2002).
53. Pollock, P. M. *et al.* High frequency of BRAF mutations in nevi. *Nature Genet.* **33**, 19–20 (2003).
54. Papp, T. *et al.* Mutational analysis of N-ras, p53, CDKN2A (p16^{INK4a}), p14^{ARF}, CDK4, and MC1R genes in human dysplastic melanocytic naevi. *J. Med. Genet.* **40**, E14 (2003).
55. Narita, M. *et al.* Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* **113**, 703–716 (2003).
56. Soengas, M. S. & Lowe, S. W. Apoptosis and melanoma chemoresistance. *Oncogene* **22**, 3138–3151 (2003).
57. Tarhini, A. A. & Agarwala, S. S. Cutaneous melanoma: available therapy for metastatic disease. *Dermatol. Ther.* **19**, 19–25 (2006).
58. Kirkwood, J. M., Moschos, S. & Wang, W. Strategies for the development of more effective adjuvant therapy of melanoma: current and future explorations of antibodies, cytokines, vaccines, and combinations. *Clin. Cancer Res.* **12**, 2331s–2336s (2006).
59. Morgan, R. A. *et al.* Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science* **314**, 126–129 (2006).
60. Smalley, K. S. & Eisen, T. G. Farnesyl transferase inhibitor SCH66336 is cytostatic, pro-apoptotic and enhances chemosensitivity to cisplatin in melanoma cells. *Int. J. Cancer* **105**, 165–175 (2003).
61. Wilhelm, S. M. *et al.* BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res.* **64**, 7099–7109 (2004).
62. Flaherty, K. T. Chemotherapy and targeted therapy combinations in advanced melanoma. *Clin. Cancer Res.* **12**, 2366s–2370s (2006).
63. Eisen, T. *et al.* Sorafenib in advanced melanoma: a Phase II randomised discontinuation trial analysis. *Br. J. Cancer* **95**, 581–586 (2006).
64. Gray-Schopfer, V., Karasides, M., Hayward, R. & Marais, R. Tumor necrosis factor- α blocks apoptosis in melanoma cells when BRAF signaling is inhibited. *Cancer Res.* **67**, 122–129 (2007).
65. Adnane, L., Trail, P. A., Taylor, I. & Wilhelm, S. M. Sorafenib (BAY 43-9006, Nexavar), a dual-action inhibitor that targets RAF/MEK/ERK pathway in tumor cells and tyrosine kinases VEGFR/PDGFR in tumor vasculature. *Methods Enzymol.* **407**, 597–612 (2005).
66. Garnett, M. J., Rana, S., Paterson, H., Barford, D. & Marais, R. Wild-type and mutant B-RAF activate C-RAF through distinct mechanisms involving heterodimerization. *Mol. Cell* **20**, 963–969 (2005).
67. Dumaz, N. *et al.* In melanoma, RAS mutations are accompanied by switching signaling from BRAF to CRAF and disrupted cyclic AMP signaling. *Cancer Res.* **66**, 9483–9491 (2006).
68. Collisson, E. A., De, A., Suzuki, H., Gambhir, S. S. & Kolodney, M. S. Treatment of metastatic melanoma with an orally available inhibitor of the Ras–Raf–MAPK cascade. *Cancer Res.* **63**, 5669–5673 (2003).
69. Solit, D. B. *et al.* BRAF mutation predicts sensitivity to MEK inhibition. *Nature* **439**, 358–362 (2006).
70. Koo, H. M. *et al.* Apoptosis and melanogenesis in human melanoma cells induced by anthrax lethal factor inactivation of mitogen-activated protein kinase kinase. *Proc. Natl Acad. Sci. USA* **99**, 3052–3057 (2002).
71. Beinke, S., Robinson, M. J., Huginun, M. & Ley, S. C. Lipopolysaccharide activation of the TPL-2/MEK/extracellular signal-regulated kinase mitogen-activated protein kinase cascade is regulated by I κ B kinase-induced proteolysis of NF- κ B1 p105. *Mol. Cell. Biol.* **24**, 9658–9667 (2004).
72. Dancy, J. E. Therapeutic targets: MTOR and related pathways. *Cancer Biol. Ther.* **5**, 1065–1073 (2006).
73. Bush, J. A. & Li, G. The role of Bcl-2 family members in the progression of cutaneous melanoma. *Clin. Exp. Metastasis* **20**, 531–539 (2003).
74. Nishimura, E. K., Granter, S. R. & Fisher, D. E. Mechanisms of hair graying: incomplete melanocyte stem cell maintenance in the niche. *Science* **307**, 720–724 (2005).
75. Jansen, B. *et al.* bcl-2 antisense therapy chemosensitizes human melanoma in SCID mice. *Nature Med.* **4**, 232–234 (1998).
76. Bedikian, A. Y. *et al.* Bcl-2 antisense (oblimersen sodium) plus dacarbazine in patients with advanced melanoma: the Oblimersen Melanoma Study Group. *J. Clin. Oncol.* **24**, 4738–4745 (2006).
77. Del Bufalo, D., Triscuoglio, D., Scarsella, M., Zangemeister-Wittke, U. & Zupi, G. Treatment of melanoma cells with a bcl-2/bcl-xL antisense oligonucleotide induces antiangiogenic activity. *Oncogene* **22**, 8441–8447 (2003).
78. Fry, D. W. *et al.* Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts. *Mol. Cancer Ther.* **3**, 1427–1438 (2004).
79. Sharp, S. & Workman, P. Inhibitors of the HSP90 molecular chaperone: current status. *Adv. Cancer Res.* **95**, 323–348 (2006).
80. da Rocha Dias, S. *et al.* Activated B-RAF is an Hsp90 client protein that is targeted by the anticancer drug 17-allylamino-17-demethoxygeldanamycin. *Cancer Res.* **65**, 10686–10691 (2005).
81. Grbovic, O. M. *et al.* V600E B-Raf requires the Hsp90 chaperone for stability and is degraded in response to Hsp90 inhibitors. *Proc. Natl Acad. Sci. USA* **103**, 57–62 (2006).
82. Amiri, K. I., Horton, L. W., LaFleur, B. J., Sosman, J. A. & Richmond, A. Augmenting chemosensitivity of malignant melanoma tumors via proteasome inhibition: implication for bortezomib (VELCADE, PS-341) as a therapeutic agent for malignant melanoma. *Cancer Res.* **64**, 4912–4918 (2004).
83. Burke, J. R. *et al.* BMS-345541 is a highly selective inhibitor of I κ B kinase that binds at an allosteric site of the enzyme and blocks NF- κ B-dependent transcription in mice. *J. Biol. Chem.* **278**, 1450–1456 (2003).
84. Rofstad, E. K. & Halsor, E. F. Vascular endothelial growth factor, interleukin 8, platelet-derived endothelial cell growth factor, and basic fibroblast growth factor promote angiogenesis and metastasis in human melanoma xenografts. *Cancer Res.* **60**, 4932–4938 (2000).
85. Tucker, G. C. Integrins: molecular targets in cancer therapy. *Curr. Oncol. Rep.* **8**, 96–103 (2006).
86. Gutheil, J. C. *et al.* Targeted antiangiogenic therapy for cancer using Vitaxin: a humanized monoclonal antibody to the integrin α v β ₃. *Clin. Cancer Res.* **6**, 3056–3061 (2000).
87. Gupta, P. B., Mani, S., Yang, J., Hartwell, K. & Weinberg, R. A. The evolving portrait of cancer metastasis. *Cold Spring Harb. Symp. Quant. Biol.* **70**, 291–297 (2005).
88. Kim, M. *et al.* Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene. *Cell* **125**, 1269–1281 (2006).
89. Straume, O. & Akslen, L. A. Alterations and prognostic significance of p16 and p53 protein expression in subgroups of cutaneous melanoma. *Int. J. Cancer* **74**, 535–539 (1997).
90. Lee, P. *et al.* ARRY-142886, a potent and selective MEK inhibitor: III) Efficacy in murine xenograft models correlates with decreased ERK phosphorylation. *Proc. Am. Assoc. Cancer Res.* **45**, 897 (2004).
91. Lyssikatos, J. *et al.* ARRY-142886, a potent and selective MEK inhibitor: I) ATP-independent inhibition results in high enzymatic and cellular selectivity. *Proc. Am. Assoc. Cancer Res.* **45**, 896-b (2004).
92. Yeh, T., Wallace, E., Lyssikatos, J. & Winkler, J. ARRY-142886, a potent and selective MEK inhibitor: II) Potency against cellular MEK leads to inhibition of cellular proliferation and induction of apoptosis in cell lines with mutant Ras or B-Raf. *Proc. Am. Assoc. Cancer Res.* **45**, 896-c–897-c (2004).
93. Morabito, A., De Maio, E., Di Maio, M., Normanno, N. & Perrone, F. Tyrosine kinase inhibitors of vascular endothelial growth factor receptors in clinical trials: current status and future directions. *Oncologist* **11**, 753–764 (2006).
94. Peterson, A. C. *et al.* Phase II study of the Flk-1 tyrosine kinase inhibitor SU5416 in advanced melanoma. *Clin. Cancer Res.* **10**, 4048–4054 (2004).
95. Margolin, K. *et al.* CCI-779 in metastatic melanoma: a phase II trial of the California Cancer Consortium. *Cancer* **104**, 1045–1048 (2005).
96. O'Donnell, A. *et al.* A phase I study of the oral mTOR inhibitor RAD001 as monotherapy to identify the optimal biologically effective dose using toxicity, pharmacokinetic (PK), and pharmacodynamic (PD) endpoints in patients with solid tumors. *Proc. Am. Soc. Clin. Oncol.* **22**, Abstr. 806 (2003).
97. End, D. W. *et al.* Characterization of the antitumor effects of the selective farnesyl protein transferase inhibitor R115777 *in vivo* and *in vitro*. *Cancer Res.* **61**, 131–137 (2001).
98. Markovic, S. N. *et al.* A phase II study of bortezomib in the treatment of metastatic malignant melanoma. *Cancer* **103**, 2584–2589 (2005).
99. Yang, J., Amiri, K. I., Burke, J. R., Schmid, J. A. & Richmond, A. BMS-345541 targets inhibitor of κ B kinase and induces apoptosis in melanoma: involvement of nuclear factor κ B and mitochondria pathways. *Clin. Cancer Res.* **12**, 950–960 (2006).
100. Yaguchi, S. *et al.* Antitumor activity of ZSTK474, a new phosphatidylinositol 3-kinase inhibitor. *J. Natl Cancer Inst.* **98**, 545–556 (2006).

Acknowledgements This work is funded by grants from The Institute of Cancer Research, Cancer Research UK, and the Association for International Cancer Research.

Author information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to R.M. (richard.marais@icr.ac.uk).

Mechanisms of sensory transduction in the skin

Ellen A. Lumpkin¹ & Michael J. Caterina²

Sensory neurons innervating the skin encode the familiar sensations of temperature, touch and pain. An explosion of progress has revealed unanticipated cellular and molecular complexity in these senses. It is now clear that perception of a single stimulus, such as heat, requires several transduction mechanisms. Conversely, a given protein may contribute to multiple senses, such as heat and touch. Recent studies have also led to the surprising insight that skin cells might transduce temperature and touch. To break the code underlying somatosensation, we must therefore understand how the skin's sensory functions are divided among signalling molecules and cell types.

The pleasant sensation of a gentle breeze or the painful experience of touching a hot stove are initiated by somatosensory neurons that innervate our skin. The peripheral terminals of these neurons, whose cell bodies are found in trigeminal and dorsal root ganglia (DRG), transduce sensory stimuli into action potentials that propagate to the central nervous system (Fig. 1).

Cutaneous sensory neurons, which are remarkably diverse, are broadly classified as A β -, A δ - or C-fibres on the basis of degree of myelination and the speed at which action potentials travel along afferent fibres¹ (Fig. 1). They can be further classified according to sensory modality¹. For example, thermoreceptors respond to warming or cooling of the skin, whereas touch receptors respond to pressure, stretch or hair movement. In addition to these neurons that respond to innocuous touch and temperatures, sensory neurons known as nociceptors initiate painful sensations. Many nociceptors are polymodal neurons that are activated by various types of sensory stimulus. The sensitivity of nociceptors to sensory stimulation can be altered by signalling pathways engaged during injury or inflammation. Sensations of pain, temperature or itch can also be evoked when endogenous or exogenous chemicals (such as histamine and menthol) activate cutaneous sensory neurons.

Here we review new findings on the cellular and molecular events that underlie somatosensation. Recent work has uncovered a number of ion channels that are candidate transducers of temperature and touch in the skin (Table 1). Attempts to nail down the roles of these ion channels *in vivo* have exposed some overlapping functions, and have suggested that additional transduction mechanisms remain to be discovered. These studies also offer the intriguing possibility that non-neuronal skin cells can directly sense touch and temperature changes. With knowledge of at least some of the factors involved in hand, the next challenge is to decode the skin's strategy for representing distinct stimuli.

Thermosensation

Mammals can discriminate temperatures ranging from extreme cold (about -10 °C) to extreme heat (about 60 °C). Different temperatures produce subjectively distinct psychophysical perceptions and objectively distinct behavioural responses. Correspondingly, different subpopulations of thermosensitive C- or A δ -fibres encode skin temperature over different ranges¹. Interestingly, the thermal responsiveness of sensory neurons is recapitulated in dissociated neurons that lack their peripheral

endings, which has proved to be invaluable for discovering candidate transduction molecules². Clues to the molecular basis of thermosensation have arisen from the recent identification of temperature-activated ion channels. Almost all of these belong to the transient receptor potential (TRP) family of cation channels³. These channels, which are divided into seven subfamilies, typically have six transmembrane domains, a pore region and cytoplasmic amino and carboxy termini, and assemble as functional tetramers³. TRP channels have a bewildering array of biophysical properties and physiological functions. Intriguingly, many participate in sensory signalling³. In mammals, thermally sensitive TRPs are each tuned to a distinct temperature range and most are expressed in cutaneous sensory neurons or other cell types in skin^{4,5} (Table 1).

Heat transduction

Among the earliest proteins implicated in heat transduction was TRP vanilloid 1 (TRPV1). This protein was identified as the molecular target of capsaicin, the main pungent component of spicy peppers². When capsaicin or other related 'vanilloid' chemicals contact the skin or mucous membranes, they activate TRPV1, which is highly expressed on nociceptive A δ - and C-fibres, to evoke a sensation of pain². Remarkably, simply heating TRPV1-expressing cells (or membrane patches derived from them) to more than 42 °C is also sufficient to evoke a robust cationic current². *In vitro*, responses to heat and capsaicin are well correlated in dissociated nociceptive neurons². Genetic ablation of TRPV1 in mice eliminates capsaicin responsiveness as well as most heat-activated currents in these cells². *In vivo*, *Trpv1* disruption results in prolonged latencies of heat-evoked paw and tail withdrawal behaviour². Importantly, however, these behavioural deficits are only partial, and are observed only at relatively high temperatures (above 50 °C)². These and other data^{6,7} indicate that TRPV1 is involved in, but cannot solely account for, acute thermal nociception in healthy skin. By contrast, TRPV1 seems to be a major contributor to the enhanced thermal responsiveness observed after cutaneous inflammation^{2,8}. Thus, the importance of TRPV1 to thermal nociception varies according to context.

Another heat-gated TRP channel, TRPV2, is strongly expressed in a population of somatosensory neurons with characteristics of A δ -nociceptors, but is also expressed in other neuronal and non-neuronal cells². Upon heterologous expression, TRPV2 is activated by very high temperatures (above 52 °C), a pattern also observed in some A δ -fibres *in vivo*

¹Departments of Neuroscience, Molecular Physiology & Biophysics and Molecular & Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

²Departments of Biological Chemistry and Neuroscience, and the Center for Sensory Biology, The Johns Hopkins School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.

and a subset of heat-sensitive DRG neurons *in vitro*². These properties suggest a role for TRPV2 in the transduction of painfully hot temperatures, although supporting evidence *in vivo* has yet to emerge.

Two other TRPV subfamily members, TRPV3 and TRPV4, can also be activated by heat, as well as by chemical and osmotic stimuli^{4,5}. TRPV4 exhibits an apparent threshold of about 27–34 °C, whereas that of TRPV3 is about 32–39 °C. These traits, combined with their expression in skin, make TRPV3 and TRPV4 candidate participants in the perception of warmth. As discussed below, the most prominent cutaneous expression of these channels is in epithelial cells rather than neurons. Support for roles for TRPV3 and TRPV4 in thermosensation has come from behavioural analyses of *Trpv3*- and *Trpv4*-null mutant mice. Both mutants exhibit abnormal thermal-selection behaviour when presented with a range of warm but non-painful floor temperatures. Furthermore, in pain-related thermal withdrawal assays, *Trpv3*-null mice show an acute thermosensation phenotype similar to that of *Trpv1*-null mice, with longer withdrawal latencies at temperatures above 50 °C. By contrast, *Trpv4*-null mice exhibit a slight increase in withdrawal latency only at 45–46 °C. Together, these findings argue in favour of differential but possibly overlapping roles for TRPV1–4 in heat perception.

Three additional TRP channels, TRPM2, TRPM4 and TRPM5, can also be activated by warm temperatures^{4,9}; however, evidence for their expression in skin is lacking.

Cold transduction

After the identification of TRPV channels as candidate heat transducers, a key question was whether cold transduction might also occur through TRP channels. This was quickly answered with the identification of TRPM8, an ion channel activated by either modest cooling from normal skin temperature (about 32 °C) to temperatures below about 30 °C, or menthol and other chemicals that produce a cooling sensation^{10,11}. In addition, TRPM8 is expressed almost exclusively in a subpopulation of C-fibres^{10,11}. These features make TRPM8 an excellent candidate cold receptor, although this hypothesis remains to be tested *in vivo*.

Some cold-responsive dissociated sensory neurons respond only to temperatures below 20 °C. One possible explanation for this behaviour has come from the characterization of TRPA1 (ref. 12), a member of the TRP ankyrin subfamily that is expressed in a subset of TRPV1-positive nociceptors^{12,13}. Ankyrin repeats — structural motifs found on many proteins, including some TRP channels — are thought to participate in protein–protein interactions. TRPA family members contain a relatively

large number of these repeats (TRPA1, for example, has 18). TRPA1 can be activated by a host of pungent chemicals containing allyl and reactive sulphur groups, such as those found in mustard oil or garlic, and by other irritants, such as acrolein^{14,15}. Some investigators have also reported that TRPA1 can be activated by the cold at less than 18 °C (refs 12, 16), prompting the suggestion that TRPA1 contributes to TRPM8-independent cold transduction.

Two *in vivo* studies have provided some support for this idea by reporting deficits in acute cold transduction or inflammation/injury-induced cold hypersensitivity in *Trpa1*^{−/−} mice¹⁷ and in rats treated with TRPA1-antisense oligonucleotides¹⁸. However, other studies have questioned these findings. First, some investigators have been unable to reproduce cold activation of heterologously expressed TRPA1 (refs 14, 19). Second, in a separate study, dissociated sensory neurons from another *Trpa1*-disrupted mouse line exhibited no discernable deficits in cold transduction, and the mice showed no differences in behavioural responses to cold temperatures²⁰. The basis for the discrepancies between these studies is unclear; however, the examination of different types of cold hypersensitivity, the use of male versus female mice, and differential strategies for gene disruption or knockdown might all be contributing factors. A definitive evaluation of the contributions of TRPA1 to cold transduction must therefore await further studies.

The involvement of TRP channels in thermosensation is not confined to mammals. In *Drosophila*, four different TRPA subfamily channels participate in thermally evoked behaviours⁴ (Table 1). As in mammals, the functions of these channels seem to segregate over different temperature ranges and result in different outputs^{4,5}.

Non-TRP ion channels might also participate in cutaneous cold signalling. For example, whether or not cold exposure triggers the firing of sensory neurons might be influenced by potassium channels, some of which exhibit strong temperature sensitivity²¹.

Mechanotransduction

A diversity of mechanosensitive neurons innervate the skin¹ (Fig. 1). Many have complex endings, some of which are associated with specialized cells in the skin. Light touch is mediated predominantly by A β afferents with low mechanical thresholds. The perception of painful touch is initiated by high-threshold C- and A δ -nociceptors that can be polymodal or solely mechanoreceptive. C-fibres sensitive to gentle touch have been described in several species, including humans, in whom they have been proposed to contribute to social interactions such as maternal bonding²².

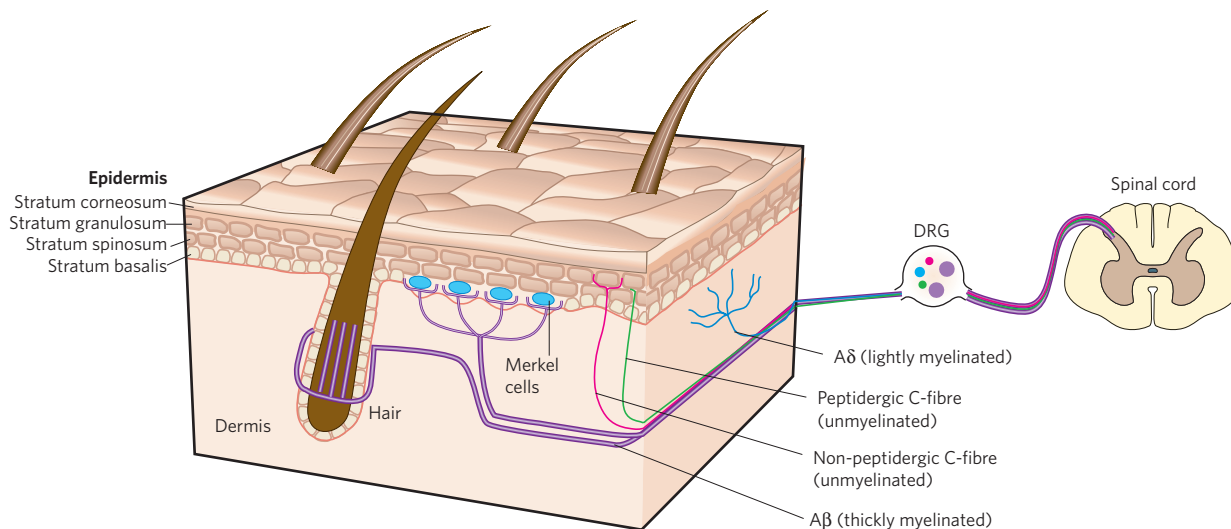


Figure 1 | Diversity of somatosensory neurons in the skin. The skin is innervated by somatosensory neurons that project to the spinal cord. A β -fibres, such as those that innervate Merkel cells and those around hair shafts, are thought to be touch receptors. A δ -fibres and C-fibres

include thermoreceptors and nociceptors. A δ -fibres terminate in the dermis. Peptidergic and non-peptidergic C-fibres terminate in different epidermal layers⁵⁹ and have different projection patterns to the spinal cord⁶⁰.

Despite intensive efforts to discover molecules that initiate touch sensation in mammals, the transduction mechanisms are largely unknown. The conceptual framework guiding these efforts posits that transduction channels are directly activated by mechanical stimuli²³. Indeed, mechanical gating might be a general model for ion channels²⁴. Stretch-sensitive channels are gated by forces in the membrane bilayer²³ (Fig. 2a). These channels can be activated experimentally by applying pressure to a membrane patch to deform the bilayer. They are found in a broad range of bacterial and eukaryotic cell types and fall into structurally distinct ion-channel families²³.

An alternative direct-gating model proposes that channels are tethered to the cytoskeleton or extracellular matrix, and that tension between these linkages controls channel gating (Fig. 2b). It is not clear whether tethers couple directly to channel domains or whether they modulate membrane forces around a stretch-sensitive channel. The tether model emerged from biophysical studies of auditory and vestibular hair cells²⁵; however, genetic screens for mechanosensory mutants in *Drosophila* and *Caenorhabditis elegans* have identified a plethora of molecular candidates that fit this model²⁶.

A third possibility is that transduction channels are coupled to mechanically sensitive proteins through signalling intermediates (Fig. 2c). For example, in *C. elegans* polymodal sensory neurons, transduction-channel activation might require the production of lipid metabolites²⁷, as described below. One limitation of indirect mechanisms is that they are intrinsically slower than direct mechanical gating. In fact, direct gating was first proposed for hair cells because of their remarkable transduction speed (about 40 μ s)²⁸. Similar arguments have been made for *C. elegans* body touch neurons²⁹ and *Drosophila* bristles³⁰; however, their reported

latencies (about 200–500 μ s) are similar to the delay between presynaptic calcium entry and postsynaptic responses at room temperature (200–600 μ s)³¹. Because this delay is sufficient to accommodate vesicle fusion, transmitter diffusion and activation of ligand-gated ion channels, this timescale alone cannot rule out indirect coupling models.

DEG/ENaC channels

The molecular basis of touch has been most extensively characterized in *C. elegans* body touch neurons²⁶. Electrophysiology²⁹ and *in vivo* imaging³² have provided direct evidence that the transduction channel is a complex of the degenerin/epithelial Na⁺ channel (DEG/ENaC) subunits MEC-4 and MEC-10, and two accessory subunits (MEC-2 and MEC-6). Although touch-evoked behaviours depend on specialized microtubules and extracellular proteins, these specialized microtubules are not essential for transduction-channel gating²⁹.

On the basis of striking *mec* phenotypes, roles for *mec*-related molecules in mammalian touch reception have been tested. Mutant mice lacking a *mec*-2-related protein, stomatin-like protein 3 (SLP3), show a marked loss of touch sensitivity *in vitro* and *in vivo*³³. Thus, SLP3 seems to be essential for mechanotransduction in a subset of cutaneous touch receptors. By contrast, similar studies indicate that three mammalian DEG/ENaC isoforms (Table 1) — the acid-sensing ion channels (ASICs) — might not have direct roles in mechanotransduction, and instead might modulate sensory signalling. In the skin, ASIC2 and 3 localize to the peripheral terminals of putative touch receptors; however, genetic disruption of ASIC subunits alters only modestly the touch-evoked responses of a few cutaneous afferent subtypes³⁴. Importantly, mechanical thresholds are not affected. Unexpectedly, mice expressing a

Table 1 | Proposed mechanosensory and thermosensory transduction channels

Identity	Family	Proposed physical modality	Additional activators	Temperature range	Sensory neuron or skin expression	Species
TRPA1	TRPA	Thermal, mechanical	Isothiocyanates, Ca ²⁺ , icilin	<18 °C	C-fibres	Mammals
Painless	TRPA	Thermal, mechanical	Isothiocyanates	n.a.	Multidendritic neurons	<i>Drosophila</i>
TRPA	TRPA	Thermal	None known	>24–29 °C	Subset of central neurons	<i>Drosophila</i>
Pyrexia-PA	TRPA	Thermal	None known	>40 °C	Multidendritic and other neurons	<i>Drosophila</i>
Pyrexia-PB	TRPA	Thermal	None known	>37 °C	Multidendritic and other neurons	<i>Drosophila</i>
TRPC1	TRPC	Mechanical	Receptor-operated, store-operated?	n.a.	Mechanosensory neurons	<i>Xenopus</i> , mammals
TRPM8	TRPM	Thermal	Menthol, icilin	<28 °C	C-fibres	Mammals
TRPN1	TRPN	Mechanical (audition)	None known	n.a.	Hair cells	<i>Danio rerio</i> , <i>Xenopus</i>
NOMPC	TRPN	Mechanical (audition, touch, proprioception)	None known	n.a.	Chordotonal organs, bristles	<i>Drosophila</i>
TRP-4	TRPN	Mechanical (touch, proprioception)	None known	n.a.	Ciliated sensory neurons and interneurons	<i>C. elegans</i>
TRPV1	TRPV	Thermal, osmotic	Capsaicin, protons, endocannabinoids, diphenyl compounds	>42 °C	C- and A δ -fibres, keratinocytes	Mammals
TRPV2	TRPV	Thermal, osmotic, mechanical	Diphenyl compounds	>52 °C	A δ - and A β -fibres, immune cells	Mammals
TRPV3	TRPV	Thermal	Camphor, carvacrol, diphenyl compounds	>34–39 °C	Keratinocytes, C-fibres	Mammals
TRPV4	TRPV	Thermal, osmotic	PUFAs, 4aPDD, epoxyeicosatrienoic acids	>27–34 °C	Keratinocytes, Merkel cells, A δ - and C-fibres	Mammals
OSM-9/OCR-2	TRPV	Thermal, osmotic	PUFAs, G-protein-coupled receptors	n.a.	Polymodal and chemosensory neurons	<i>C. elegans</i>
NAN/IAV	TRPV	Mechanical (audition, proprioception)	Osmotic	n.a.	Chordotonal neurons	<i>Drosophila</i>
ASIC1	DEG/ENaC	Mechanical (touch)	Protons	n.a.	A δ -, A β - and C-fibres	Mammals
ASIC2	DEG/ENaC	Mechanical (touch)	Protons	n.a.	A δ - and A β -fibres	Mammals
ASIC3	DEG/ENaC	Mechanical (touch, nociception)	Protons	n.a.	A δ - and A β -fibres	Mammals
MEC-4/MEC-10	DEG/ENaC	Mechanical (touch)	None known	n.a.	Body touch neurons	<i>C. elegans</i>
TREK-1	2P K ⁺ channel	Thermal, mechanical	Lipids, protons	n.a.	A δ - and C-fibres, A β -fibres?	Mammals

dominant-negative *Asic3* mutation are hypersensitive to acute mechanical and chemical stimuli³⁵.

TRPN and TRPA channels

Like the *mec* genes, *Drosophila* NOMPC (TRPN1) was identified as a candidate transduction channel in a genetic screen for touch-insensitive animals³⁰. TRPN channels have unusually large N-terminal domains with 29 ankryin repeats. *Drosophila* *nompC* mutants show defects in hearing, touch and proprioception. Moreover, the *C. elegans* TRPN1 homologue *trp-4* is expressed in mechanosensory neurons^{30,36} and has been proposed to function in proprioception on the basis of the conspicuous locomotory defects of *trp-4* mutants³⁶.

In *Drosophila* auditory antennae and vertebrate hair cells, mechanical stimuli initiate two processes that govern auditory sensitivity: electrical excitation that triggers downstream neuronal signalling and mechanical amplification that increases sensitivity at particular sound frequencies. Interestingly, *nompC* mutations in *Drosophila* partly reduce mechanically evoked neuronal signals in bristles³⁰ and antennae³⁷ but completely eliminate mechanical amplification in antennae³⁸. Because the transduction machinery is an essential part of the amplification feedback loop, these results suggest that NOMPC directly participates in transduction but that it cannot be the only transduction channel in these neurons.

The vertebrate homologue of NOMPC, TRPN1, is expressed in mechanosensory hair cells in some fish and amphibians^{39,40} but is not present in the genomes of reptiles, birds or mammals. Hair cells contain a cluster of mechanically sensitive microvilli, known as stereocilia, and a single true cilium, the kinocilium. The latter is not required for mechanotransduction, although it is probably important for the development or stimulation of stereocilia *in vivo*. In zebrafish hair cells, RNA-interference knockdown of *trpn1* eliminates mechanically evoked potentials³⁹. Surprisingly, the channel's predominant localization to kinocilia^{39,40} indicates that it might not directly mediate transduction. Collectively, these studies point towards a phylogenetically conserved role for TRPN channels in mechanosensory signalling; however, TRPN subunits might also have roles other than transduction.

The intriguing phenotypes of TRPN mutants focused attention on TRPA1 because it is the only mammalian TRP channel with an extended ankryin domain and it is expressed in nociceptors^{12,13,19} and hair cells^{19,41}. Interestingly, one *Drosophila* *Trpa* isoform, *painless*, is required for withdrawal from harsh prodding⁴². By contrast, *Trpa1*^{-/-} mice have been reported to show small¹⁷ or no²⁰ deficits in acute touch sensitivity, indicating that TRPA1 is not essential for mechanotransduction in somatosensory neurons. Although RNA-interference-mediated knockdown suggested TRPA1 to be a promising candidate in hair-cell transduction⁴¹, *Trpa1*^{-/-} mice have normal auditory responses^{17,20} and hair-cell transduction currents¹⁷. Together, these results indicate that mammalian TRPA1 has a minor role in the acute transduction of mechanical stimuli; however, the *Trpa1*-knockout studies demonstrated that this channel is crucial for the hypersensitivity to both touch and heat that accompanies skin inflammation by mustard oil as well as the heat hypersensitivity caused by bradykinin. Thus, TRPA1 may be an excellent target for new therapeutics for pain hypersensitivity.

TRPV channels

The first TRP channel shown to participate in mechanosensation was the *C. elegans* TRPV channel OSM-9 (ref. 43). In the polymodal sensory neuron ASH, OSM-9 and another TRPV isoform, OCR-2, co-localize to sensory cilia, which is consistent with the idea that they form heteromeric transduction channels. Like other TRPV channels, OSM-9/OCR-2 channels are thought to be polymodal, because mutations in each subunit disrupt avoidance of hypertonic, mechanical and chemical stimuli.

Although these stimuli probably activate OSM-9/OCR-2 channels through distinct signalling cascades, each cascade might involve polyunsaturated fatty acids (PUFAs)⁴³. In fact, modulation by lipid metabolites is a feature of many TRP channels, including *Drosophila* *trp*³. In *C. elegans*, mutations that interfere with PUFA synthesis have been shown to com-

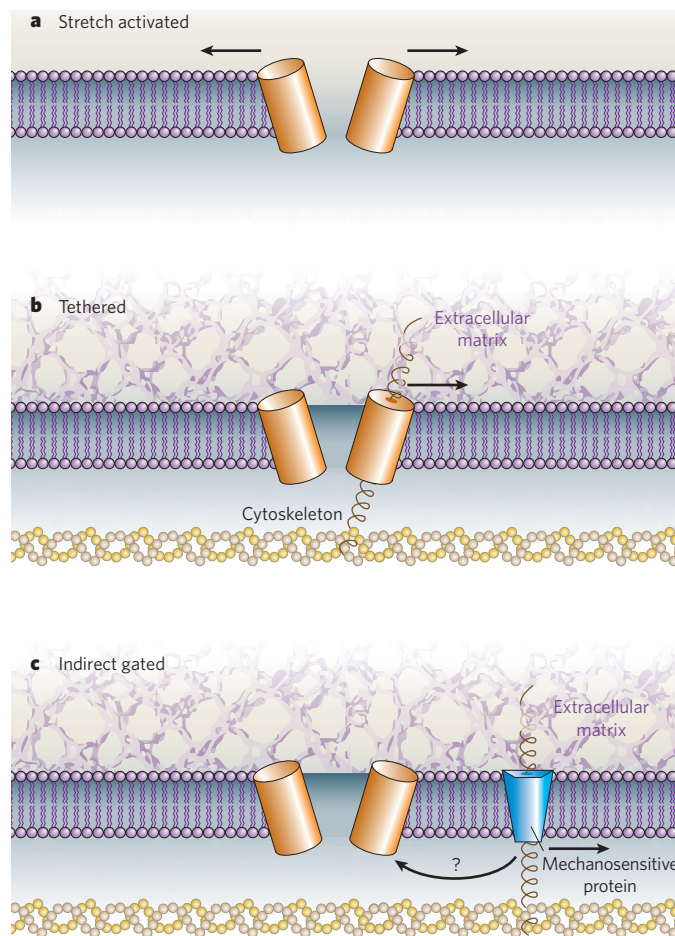


Figure 2 | Gating models of mechanotransduction channels. **a**, Stretch-activated ion channels open when forces (horizontal arrows) in the lipid bilayer change, for example, owing to alterations in bilayer tension or curvature. **b**, In sensory cells, mechanically gated channels are proposed to require links to extracellular or cytoskeletal proteins. Displacements that change the tension on these links open the channel. These links could directly transmit force to the channel protein (as depicted) or could control the membrane forces around stretch-sensitive channels. **c**, Another possibility is that a mechanosensitive protein regulates ion-channel opening through a signalling intermediate. Such hypothetical transduction proteins might require tethers (as depicted) or might respond to changes in the lipid bilayer.

promise all three ASH-mediated sensory modalities as well as olfactory responses²⁷. Conversely, exogenously applied PUFAs rapidly activate ASH and elicit avoidance behaviours in an *osm-9*-dependent manner. Together, these data suggest that PUFAs are produced through sensory transduction cascades to activate OSM-9; however, it is also possible that these lipids serve as channel modulators or provide a lipid environment that is permissive for channel function.

The two *Drosophila* TRPV isoforms, *nanchung* (*nan*) and *inactive* (*iav*), are essential for sensory signalling because mutations in either gene abolish sound-evoked neuronal activity in auditory antennae^{44,45}. Moreover, NAN and IAV are activated by hypotonic challenge in heterologous cell types, which suggests that they can transduce mechanical stimuli. Like OSM-9/OCR-2, NAN and IAV co-localize to sensory cilia of mechanoreceptive neurons^{44,45}. Surprisingly, a green-fluorescent-protein-tagged version of IAV was found only in the proximal cilium⁴⁴, which is at odds with the hypothesis that mechanically gated channels are tethered at ciliary tips. Moreover, whereas *nompC* mutations abolish mechanical amplification in auditory antennae, *nan* and *iav* loss-of-function mutations actually increase amplification³⁸. This suggests that NAN and IAV are not part of the mechanically activated transduction complex. An alternative model that can reconcile these findings is that

mechanically sensitive proteins indirectly activate NAN/IAV channels to effect downstream signalling (Fig. 2c). One speculative model is that NAN/IAV, like OSM-9/OCR-2, are activated by fatty acid metabolites, although such signalling might not be fast enough to account for fly auditory transduction^{37,44}.

Along with its role in thermosensation, mammalian TRPV4 has been proposed to function in mechanotransduction and osmosensation⁴⁶. Similarly to NAN and IAV, heterologously expressed TRPV4 can be activated by hypotonic solutions^{47,48}. As with OSM-9, the osmotic activation of TRPV4 requires fatty acid metabolites⁴⁹. Moreover, TRPV4 can complement some of the sensory defects of *C. elegans* *osm-9* mutants⁴⁶. Although its expression in large-diameter sensory neurons and Merkel cells⁴⁷ is consistent with a role in cutaneous touch, disrupting *Trpv4* expression in mice has only modest effects on acute mechanosensory thresholds^{50,51}. TRPV4 function might instead be important for inflammation-induced mechanical hypersensitivity and nociceptive responses to hypotonic solutions⁵².

Two additional mammalian TRPV subunits have been implicated in mechanosensory signalling. TRPV1 is dispensable for cutaneous mechanosensation; however, this channel is required for normal stretch-evoked reflexes in the bladder⁵³ and for osmosensation in hypothalamic neurons⁵⁴. Mammalian TRPV2 is a candidate mechanotransduction channel because it can be activated by hypotonic and stretch stimuli *in vitro*⁵⁵ and it is expressed in large-diameter somatosensory neurons².

Stretch-sensitive channels

Stretch-sensitive ion channels that are expressed in sensory neurons have also been proposed to participate in mammalian mechanotransduction. For example, the canonical TRP channel TRPC1, which is activated by membrane stretch in *Xenopus* oocytes⁵⁶, is broadly expressed in mammalian cells, including in somatosensory neurons. In addition, the stretch-sensitive two-pore potassium channel TREK-1 seems to have a key role in acute touch: *Trek-1*^{-/-} mice show markedly increased sensitivity to low-threshold mechanical stimuli, but normal sensitivity to acute heat and noxious pressure⁵⁷. These results highlight the fact that many types of channel must work in concert to control the sensitivity of touch and pain receptors.

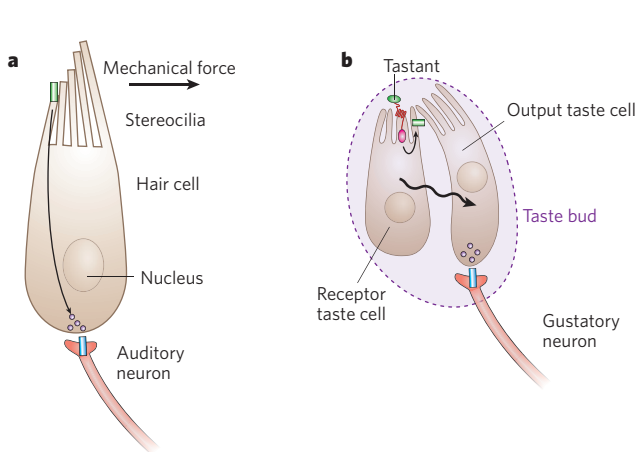


Figure 3 | Sensory transduction by epithelial cells. **a**, Hair cell of the inner ear. Mechanical deflection of stereocilia opens non-selective cation channels, which depolarize the cell to increase the rate of glutamate release onto synaptically connected auditory afferents. **b**, Taste bud epithelium. Chemical tastants activate a G-protein-coupled receptor-ion channel pathway in a 'receptor' epithelial cell, which communicates through paracrine signalling (wavy line) to an adjacent 'output' epithelial cell. The output cell releases neurotransmitter onto a synaptically connected taste afferent⁵⁸. **c**, Proposed models for keratinocyte and Merkel cell involvement

Roles for epidermal cells in somatosensation

Somatosensory transduction is generally thought to occur in the terminals of sensory neurons. Skin cells are often viewed as bystanders in this process, providing, at most, physical and trophic support to those nerve terminals. An emerging alternative view is that non-neuronal cells function as primary transducers of some physical and chemical stimuli and that these cells can, in turn, communicate to neighbouring somatosensory afferents.

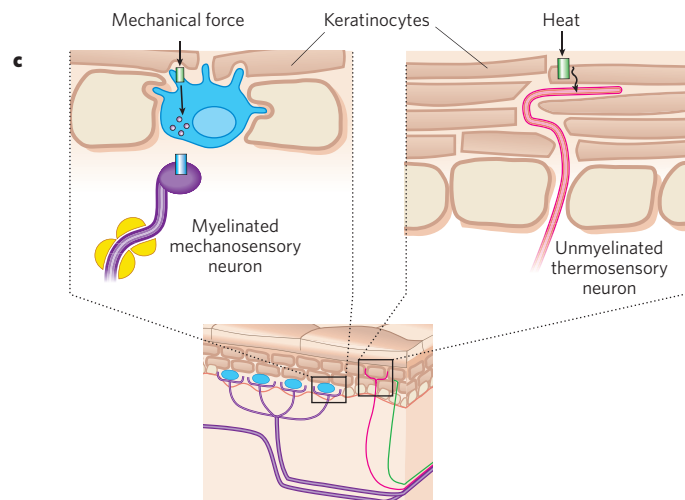
Epithelial cells as sensory receptor cells

Two epithelial cell types are well established as mediators of sensory transduction. First, hair cells carry out the mechanotransduction that underlies auditory and vestibular function. In these cells, displacement opens transduction channels near the tips of stereocilia, causing membrane depolarization and consequent release of neurotransmitter (almost certainly glutamate) from the hair cell's synapses onto afferent auditory neurons²⁵ (Fig. 3a). Second, taste perception begins with chemical transduction events that occur in specialized epithelial cells in the taste buds (Fig. 3b). The signal by which taste cells communicate with afferent neurons has not been firmly established, but ATP and serotonin are strong candidates⁵⁸. Furthermore, there is evidence for a signal relay of sorts by which tastants are initially detected by a 'receptor' epithelial cell that communicates with a neighbouring 'output' epithelial cell that, in turn, signals to the sensory afferent⁵⁸ (Fig. 3b).

Keratinocytes

Keratinocytes, the predominant cell type in the epidermis, proliferate from a basal layer located at the dermal-epidermal border and, through a coordinated programme of differentiation and apical migration, form a well-organized stratified epithelium (Fig. 1). Although much attention has been focused on the chemical and mechanical barrier functions of keratinocytes, physical protection seems to be just one role of these cells.

Three lines of evidence suggest that keratinocytes participate in the detection of physical and chemical stimuli. First, localization studies involving genetically labelled neuronal subpopulations have revealed at least two populations of sensory afferent fibres that ramify within distinct epidermal layers⁵⁹ (Fig. 1). Sensory neurons that express proinflammatory neuropeptides such as calcitonin gene-related polypeptide



in cutaneous thermotransduction and mechanotransduction. Touch activates an unknown transduction mechanism in Merkel cells (blue, left), which are synaptically connected to cutaneous A β afferents. The myelin sheath of the A β afferent is lost as it approaches a Merkel cell. Projections from the Merkel-cell surface represent microvilli. Heat activates TRPV3 and TRPV4 ion channels (green, right) expressed in keratinocytes. Paracrine signalling from keratinocytes to adjacent afferents (wavy line) results in neuronal activation. For simplicity, direct thermotransduction and mechanotransduction by cutaneous afferents are not shown.

(CGRP) ascend through the basal epidermal layer to terminate in the next layer, the stratum spinosum. By contrast, a population of nonpeptidergic neurons that express a specific G-protein-coupled receptor, MrgD, terminate more superficially, in the stratum granulosum. These two neuronal populations also exhibit distinct projection patterns to the spinal cord and have apparently distinct sensory functions⁶⁰. Although classical synaptic structures between keratinocytes and sensory nerve terminals have not been described, the proximity of these cell types and even close membrane–membrane apposition^{61,62} provides ample opportunity for rapid paracrine communication.

A second line of evidence implicating keratinocytes in sensory signalling comes from the fact that these cells secrete numerous chemical substances capable of modulating, activating or inhibiting sensory neurons. Examples include neurotrophins, ATP, β -endorphin, interleukins and endothelin-1 (refs 63–66). Interestingly, different layers of the epidermis can release substances with different effects. Whereas superficial keratinocytes can release antinociceptive molecules such as β -endorphin, stimulation of deeper epidermal keratinocytes results in the release of pro-nociceptive endothelin-1 (ref. 67). The presence of such potentially antagonistic systems in the epidermis hints at a capacity for signal filtration or processing upstream of the nervous system.

A third fact that supports the involvement of keratinocyte function in acute sensory signalling is that they express various receptors that have been implicated in pain or temperature sensation. Some of these are receptors for the chemical substances mentioned above. In these cases, it is difficult to distinguish whether the receptors facilitate interkeratinocyte signalling, keratinocyte–neuron signalling, or both. For example, mechanically stimulating keratinocytes *in vitro* causes ATP release and signalling through the metabotropic ATP receptor P2Y2 (ref. 66). P2Y2 activation mobilizes the release of intracellular Ca^{2+} stores, which, in turn, evokes the release of still more ATP from the stimulated keratinocytes. The result is an intercellular relay that spreads across the culture. If sensory neurons are co-cultured with keratinocytes stimulated in this way, the neurons exhibit a delayed activation through their own purinergic receptors⁶⁶. Although this sequence of events has not been established *in vivo*, it provides a plausible picture of how a physical stimulus could result in sensory neuron excitation.

Recent data also indicate that keratinocytes might transduce thermal stimuli. For example, the warm-activated ion channels TRPV3 and TRPV4 are more readily detectable in keratinocytes than in sensory neurons^{4–6}. Currents mediated by these two types of channel can be evoked by heat stimulation of cultured keratinocytes and distinguished on the basis of their kinetic profiles^{4,5}. These responses can be augmented by agonists of TRPV3 or TRPV4 (camphor or 4- α -phorbol 12,13-didecanoate (4aPDD), respectively) and are eliminated by *Trpv3* or *Trpv4* disruption^{4,5}.

The expression of functional TRPV3 and TRPV4 in keratinocytes, coupled with the behavioural thermosensory defects in *Trpv3*- or *Trpv4*-null mutant mice provides a strong circumstantial case for the contribution of the keratinocyte-expressed channels in heat sensation. Whether — and, if so, how — stimulation of these channels results in neuronal activation has not yet been established, however. One recent study demonstrated that chemical activation of TRPV3 in keratinocytes cultured from tongue epithelium caused interleukin-1 α release⁶⁵, providing validation of the idea that TRPV channel activation in keratinocytes stimulates the release of bioactive substances.

TRPV1 expression in keratinocytes has also been reported, and capsaicin can evoke the release of interleukin-8 from immortalized human keratinocytes⁵. Similarly, in the urinary bladder TRPV1 is expressed both in sensory neurons and in the urothelial cells that form the bladder luminal epithelium. In fact, TRPV1 is essential for the stretch-evoked release of ATP from the urothelium, a process that triggers afferent neuron activation and reflex contraction of the bladder⁵.

Despite these findings, direct evidence for acute keratinocyte to sensory neuron signalling is lacking. Furthermore, ATP receptors and TRPV channels in keratinocytes almost certainly have other functions. For example, extracellular ATP is required for the barrier function of

keratinocytes. In addition, abnormalities in hair integrity⁶⁸ or hair follicle cycling⁶⁹ have been detected in *Trpv3*- and *Trpv1*-null mice, respectively, as well as in rodents expressing a mutated form of *Trpv3* (ref. 70).

Merkel cells

Merkel cell–neurite complexes, which mediate a subset of slowly adapting responses to touch, are important for distinguishing shape, form and texture⁷¹. These complexes are made up of Merkel cells in close association with A β -afferent terminals (Fig. 3c). They are found in the basal epidermal layer of touch-sensitive areas of the skin, including glabrous skin, whisker follicles and touch domes.

Parallels identified between Merkel cells and hair cells have fuelled speculation that Merkel cells are sensory cells that transduce touch and then communicate with afferents by synaptic transmission⁷¹ (Fig. 3c). First, both epithelial cell types seem to make synaptic contacts with sensory terminals. Second, Merkel cells contain elongated microvilli, reminiscent of the hair cell's stereocilia. In fact, Merkel cells express an espin isoform that is found only in the microvilli of sensory cells, such as hair cells, taste receptor cells and vomeronasal neurons⁷². Finally, Merkel cells and hair cells express some of the same developmental transcription factors, including mammalian atonal homologue 1 (ref. 73) and growth-factor independent 1 (ref. 74). Although these similarities are intriguing, experiments designed to determine whether Merkel cells are required for touch sensitivity have produced contradictory results⁷¹ and so far there is no direct evidence demonstrating that Merkel cells are activated by touch.

The idea that Merkel cells are presynaptic is supported by morphological and molecular studies⁷¹; however, functional studies that tested the role of synaptic transmission in touch have produced conflicting results^{71,75–77}. Merkel cells contain dense-core vesicles that resemble neurosecretory vesicles. Moreover, Merkel cells have membrane densities similar to those at synaptic active zones and they express piccolo, a pre-synaptic active-zone protein⁷⁸. Because Merkel cells express molecules necessary for synaptic glutamate release and neuropeptide production⁷⁸, an important open question is whether Merkel cell synapses are excitatory or whether they send modulatory signals to regulate touch-sensitive neurons.

Another challenge to the simple model that Merkel cells are sensory cells is that recent immunohistochemical studies have localized a number of neurotransmitter receptors to Merkel cells but not to associated sensory neurons^{76,79}. These data suggest that neurotransmitters might be released from sensory neurons or keratinocytes to signal to Merkel cells or perhaps that Merkel-cell signalling is autocrine.

Future directions

A flurry of recent work has indicated that many ion channels are in the right place and have the right properties to participate in the skin's sensory functions. These advances have uncovered recurring themes but have also provided a tantalizing glimpse of the molecular complexity of somatosensory transduction. A key challenge now is to explain the molecular logic of touch, temperature sensation and nociception.

One unifying principle is the participation of TRP channels in somatosensation. The extent to which these channels exhibit specialized versus overlapping functions remains to be clarified. Their heterogeneity might be further increased by heteromultimerization among TRP subunits or splice variants. Moreover, changes in channel expression, sensitivity or interactions might contribute to alterations in sensory perception that accompany tissue inflammation.

A second recurring theme is the polymodality of putative transduction channels, which raises the possibility that sensory integration begins at the first step of signalling. Whether and how stimuli converge to modulate ion channels *in vivo* are crucial open questions. To answer them, we must first understand how physical and chemical stimuli control channel gating. For example, an important step is to determine how force opens mechanotransduction channels, such as the MEC-4/MEC-10 complex. Structure and function studies have begun to define TRP channel domains required for activation by chemical stimuli^{80,81}.

Moreover, thermodynamics-based efforts have yielded some insight into thermosensitivity⁸². Nonetheless, our understanding of polymodal channel gating remains rudimentary and requires higher-resolution structural and functional studies.

A theme revealed by invertebrate systems is the molecular heterogeneity of mechanotransduction. Although vertebrate DEG/ENaC and TRP channels are currently under the greatest scrutiny as candidate transduction channels, the subtle mechanosensory defects in knockout mice suggest that there are other candidates still to be identified. It is possible that structurally unrelated mechanotransduction channels have yet to be discovered, as was highlighted by the recent finding that Ca²⁺-release-activated Ca²⁺ channels belong to a previously unrecognized ion channel family⁸³. We also need new strategies for evaluating candidate mechanotransduction channels. For example, techniques for measuring transduction currents in the skin are essential for determining whether a candidate acts directly in transduction or in signal propagation.

A final emerging concept is the role of epithelial cells in sensory transduction. A fundamental open question is whether skin cells truly act as first-line transducers of physical stimuli, or whether their expression of TRP channels reflects modulatory or non-sensory functions. If epithelial cells do mediate sensory transduction, we must discover mechanisms of signal relay and specificity between epithelial cells and neurons. For example, a molecule released from heat-activated keratinocytes must excite the correct sensory neuron subtypes and must be short-lived enough to allow discrimination of acute changes in skin temperature. For Merkel cells, the synaptic junctions with A β -fibres provide a mechanism for specificity and speed; however, the direction and type of signalling between Merkel cells and sensory neurons are open questions. ■

- Gardner, E. P., Martin, J. H. & Jessell, T. M. in *Principles of Neural Science* (eds Kandel, E. R., Schwartz, J. H. & Jessell, T. M.) 430–449 (Oxford Univ. Press, New York, 2000).
- Caterina, M. J. & Julius, D. The vanilloid receptor: a molecular gateway to the pain pathway. *Annu. Rev. Neurosci.* **24**, 487–517 (2001).
- Ramsey, I. S., Delling, M. & Clapham, D. E. An introduction to TRP channels. *Annu. Rev. Physiol.* **68**, 619–647 (2006).
- Dhaka, A., Viswanath, V. & Patapoutian, A. TRP ion channels and temperature sensation. *Annu. Rev. Neurosci.* **29**, 135–161 (2006).
- Lee, H. & Caterina, M. J. TRPV channels as thermosensory receptors in epithelial cells. *Pflügers Arch.* **451**, 160–167 (2005).
- Zimmermann, K. et al. The TRPV1/2/3 activator 2-aminoethoxydiphenyl borate sensitizes native nociceptive neurons to heat in wildtype but not TRPV1 deficient mice. *Neuroscience* **135**, 1277–1284 (2005).
- Woodbury, C. J. et al. Nociceptors lacking TRPV1 and TRPV2 have normal heat responses. *J. Neurosci.* **24**, 6410–6415 (2004).
- Szallasi, A., Cruz, F. & Geppetti, P. TRPV1: a therapeutic target for novel analgesic drugs? *Trends Mol. Med.* **12**, 545–554 (2006).
- Togashi, K. et al. TRPM2 activation by cyclic ADP-ribose at body temperature is involved in insulin secretion. *EMBO J.* **25**, 1804–1815 (2006).
- Peier, A. M. et al. A TRP channel that senses cold stimuli and menthol. *Cell* **108**, 705–715 (2002).
- McKemy, D. D., Neuhauser, W. M. & Julius, D. Identification of a cold receptor reveals a general role for TRP channels in thermosensation. *Nature* **416**, 52–58 (2002).
- Story, G. M. et al. ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* **112**, 819–829 (2003).
- Bautista, D. M. et al. Pungent products from garlic activate the sensory ion channel TRPA1. *Proc. Natl Acad. Sci. USA* **102**, 12248–12252 (2005).
- Jordt, S. E. et al. Mustard oils and cannabinoids excite sensory nerve fibres through the TRP channel ANKTM1. *Nature* **427**, 260–265 (2004).
- Bandell, M. et al. Noxious cold ion channel TRPA1 is activated by pungent compounds and bradykinin. *Neuron* **41**, 849–857 (2004).
- Reid, G. ThermoTRP channels and cold sensing: what are they really up to? *Pflügers Arch.* **451**, 250–263 (2005).
- Kwan, K. Y. et al. TRPA1 contributes to cold, mechanical, and chemical nociception but is not essential for hair-cell transduction. *Neuron* **50**, 277–289 (2006).
- Obata, K. et al. TRPA1 induced in sensory neurons contributes to cold hyperalgesia after inflammation and nerve injury. *J. Clin. Invest.* **115**, 2393–2401 (2005).
- Nagata, K., Duggan, A., Kumar, G. & Garcia-Anoveros, J. Nociceptor and hair cell transducer properties of TRPA1, a channel for pain and hearing. *J. Neurosci.* **25**, 4052–4061 (2005).
- Bautista, D. M. et al. TRPA1 mediates the inflammatory actions of environmental irritants and proalgesic agents. *Cell* **124**, 1269–1282 (2006).
- Viana, F., de la Pena, E. & Belmonte, C. Specificity of cold thermotransduction is determined by differential ionic channel expression. *Nature Neurosci.* **5**, 254–260 (2002).
- Olausson, H. et al. Unmyelinated tactile afferents signal touch and project to insular cortex. *Nature Neurosci.* **5**, 900–904 (2002).
- Kung, C. A possible unifying principle for mechanosensation. *Nature* **436**, 647–654 (2005).
- Jiang, Y. et al. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
- LeMasurier, M. & Gillespie, P. G. Hair-cell mechanotransduction and cochlear amplification. *Neuron* **48**, 403–415 (2005).
- Syntichaki, P. & Tavernarakis, N. Genetic models of mechanotransduction: the nematode *Caenorhabditis elegans*. *Physiol. Rev.* **84**, 1097–1153 (2004).
- Kahn-Kirby, A. H. et al. Specific polyunsaturated fatty acids drive TRPV-dependent sensory signaling in vivo. *Cell* **119**, 889–900 (2004).
- Corey, D. P. & Hudspeth, A. J. Response latency of vertebrate hair cells. *Biophys. J.* **26**, 499–506 (1979).
- O'Hagan, R., Chalfie, M. & Goodman, M. B. The MEC-4 DEG/ENaC channel of *Caenorhabditis elegans* touch receptor neurons transduces mechanical signals. *Nature Neurosci.* **8**, 43–50 (2005).
- Walker, R. G., Willingham, A. T. & Zuker, C. S. A *Drosophila* mechanosensory transduction channel. *Science* **287**, 2229–2234 (2000).
- Sabatini, B. L. & Regehr, W. G. Timing of neurotransmission at fast synapses in the mammalian brain. *Nature* **384**, 170–172 (1996).
- Suzuki, H. et al. In vivo imaging of *C. elegans* mechanosensory neurons demonstrates a specific role for the MEC-4 channel in the process of gentle touch sensation. *Neuron* **39**, 1005–1017 (2003).
- Wetzel, C. et al. A stomatin-domain protein essential for touch sensation in the mouse. *Nature* **445**, 206–209 (2007).
- Lumpkin, E. A. & Bautista, D. M. Feeling the pressure in mammalian somatosensation. *Curr. Opin. Neurobiol.* **15**, 382–388 (2005).
- Mogil, J. S. et al. Transgenic expression of a dominant-negative ASIC3 subunit leads to increased sensitivity to mechanical and inflammatory stimuli. *J. Neurosci.* **25**, 9893–9901 (2005).
- Li, W., Feng, Z., Sternberg, P. W. & Xu, X. Z. A *C. elegans* stretch receptor neuron revealed by a mechanosensitive TRP channel homologue. *Nature* **440**, 684–687 (2006).
- Eberl, D. F., Hardy, R. W. & Kernan, M. J. Genetically similar transduction mechanisms for touch and hearing in *Drosophila*. *J. Neurosci.* **20**, 5981–5988 (2000).
- Gopfert, M. C., Albert, J. T., Nadrowski, B. & Kamikouchi, A. Specification of auditory sensitivity by *Drosophila* TRP channels. *Nature Neurosci.* **9**, 999–1000 (2006).
- Sidi, S., Friedrich, R. W. & Nicolson, T. NompC TRP channel required for vertebrate sensory hair cell mechanotransduction. *Science* **301**, 96–99 (2003).
- Shin, J. B. et al. *Xenopus* TRPN1 (NOMPC) localizes to microtubule-based cilia in epithelial cells, including inner-ear hair cells. *Proc. Natl Acad. Sci. USA* **102**, 12572–12577 (2005).
- Corey, D. P. et al. TRPA1 is a candidate for the mechanosensitive transduction channel of vertebrate hair cells. *Nature* **432**, 723–730 (2004).
- Tracey, W. D., Wilson, R. I., Laurent, G. & Benzer, S. *painless*, a *Drosophila* gene essential for nociception. *Cell* **113**, 261–273 (2003).
- Kahn-Kirby, A. H. & Bargmann, C. I. TRP channels in *C. elegans*. *Annu. Rev. Physiol.* **68**, 719–736 (2006).
- Gong, Z. et al. Two interdependent TRPV channel subunits, inactive and Nanchung, mediate hearing in *Drosophila*. *J. Neurosci.* **24**, 9059–9066 (2004).
- Kim, J. et al. A TRPV family ion channel required for hearing in *Drosophila*. *Nature* **424**, 81–84 (2003).
- Liedtke, W. TRPV4 as osmosensor: a transgenic approach. *Pflügers Arch.* **451**, 176–180 (2005).
- Liedtke, W. et al. Vanilloid receptor-related osmotically activated channel (VR-OAC), a candidate vertebrate osmoreceptor. *Cell* **103**, 525–535 (2000).
- Strotmann, R., Harteneck, C., Nunnenmacher, K., Schultz, G. & Plant, T. D. OTRPC4, a nonselective cation channel that confers sensitivity to extracellular osmolarity. *Nature Cell Biol.* **2**, 695–702 (2000).
- Vriens, J. et al. Cell swelling, heat, and chemical agonists use distinct pathways for the activation of the cation channel TRPV4. *Proc. Natl Acad. Sci. USA* **101**, 396–401 (2004).
- Liedtke, W. & Friedman, J. M. Abnormal osmotic regulation in *trpv4*^{-/-} mice. *Proc. Natl Acad. Sci. USA* **100**, 13698–13703 (2003).
- Suzuki, M., Mizuno, A., Kodaira, K. & Imai, M. Impaired pressure sensation in mice lacking TRPV4. *J. Biol. Chem.* **278**, 22664–22668 (2003).
- Alessandri-Haber, N., Joseph, E., Dina, O. A., Liedtke, W. & Levine, J. D. TRPV4 mediates pain-related behavior induced by mild hypertonic stimuli in the presence of inflammatory mediator. *Pain* **118**, 70–79 (2005).
- Birder, L. A. et al. Altered urinary bladder function in mice lacking the vanilloid receptor TRPV1. *Nature Neurosci.* **5**, 856–860 (2002).
- Sharif Naeini, R., Witty, M. F., Seguela, P. & Bourque, C. W. An N-terminal variant of Trpv1 channel is required for osmosensory transduction. *Nature Neurosci.* **9**, 93–98 (2006).
- Muraki, K. et al. TRPV2 is a component of osmotically sensitive cation channels in murine aortic myocytes. *Circ. Res.* **93**, 829–838 (2003).
- Maroto, R. et al. TRPC1 forms the stretch-activated cation channel in vertebrate cells. *Nature Cell Biol.* **7**, 179–185 (2005).
- Alloui, A. et al. TREK-1, a K⁺ channel involved in polymodal pain perception. *EMBO J.* **25**, 2368–2376 (2006).
- Roper, S. D. Signaling in the chemosensory systems: cell communication in taste buds. *Cell Mol. Life Sci.* **63**, 1494–1500 (2006).
- Zylka, M. J., Rice, F. L. & Anderson, D. J. Topographically distinct epidermal nociceptive circuits revealed by axonal tracers targeted to Mrgprd. *Neuron* **45**, 17–25 (2005).
- Snider, W. D. & McMahon, S. B. Tackling pain at the source: new ideas about nociceptors. *Neuron* **20**, 629–632 (1998).
- Hilliges, M., Wang, L. & Johansson, O. Ultrastructural evidence for nerve fibers within all vital layers of the human epidermis. *J. Invest. Dermatol.* **104**, 134–137 (1995).
- Chateau, Y. & Misery, L. Connections between nerve endings and epidermal cells: are they synapses? *Exp. Dermatol.* **13**, 2–4 (2004).
- Shu, X. Q. & Mendell, L. M. Neurotrophins and hyperalgesia. *Proc. Natl Acad. Sci. USA* **96**, 7693–7696 (1999).
- Khodorova, A., Fareed, M. U., Gokin, A., Strichartz, G. R. & Davar, G. Local injection of a selective endothelin-B receptor agonist inhibits endothelin-1-induced pain-like behavior

- and excitation of nociceptors in a naloxone-sensitive manner. *J. Neurosci.* **22**, 7788–7796 (2002).
65. Xu, H., Dellling, M., Jun, J. C. & Clapham, D. E. Oregano, thyme and clove-derived flavors and skin sensitizers activate specific TRP channels. *Nature Neurosci.* **9**, 628–635 (2006).
 66. Koizumi, S. *et al.* Ca^{2+} waves in keratinocytes are transmitted to sensory neurons: the involvement of extracellular ATP and P2Y2 receptor activation. *Biochem. J.* **380**, 329–338 (2004).
 67. Khodorova, A. *et al.* Endothelin-B receptor activation triggers an endogenous analgesic cascade at sites of peripheral injury. *Nature Med.* **9**, 1055–1061 (2003).
 68. Moqrich, A. *et al.* Impaired thermosensation in mice lacking TRPV3, a heat and camphor sensor in the skin. *Science* **307**, 1468–1472 (2005).
 69. Biro, T. *et al.* Hair cycle control by vanilloid receptor-1 (TRPV1): evidence from TRPV1 knockout mice. *J. Invest. Dermatol.* **126**, 1909–1912 (2006).
 70. Asakawa, M. *et al.* Association of a mutation in TRPV3 with defective hair growth in rodents. *J. Invest. Dermatol.* **126**, 2664–2672 (2006).
 71. Halata, Z., Grim, M. & Bauman, K. I. Friedrich Sigmund Merkel and his 'Merkel cell', morphology, development, and physiology: review and new results. *Anat. Rec.* **271A**, 225–239 (2003).
 72. Sekerkova, G. *et al.* Espins are multifunctional actin cytoskeletal regulatory proteins in the microvilli of chemosensory and mechanosensory cells. *J. Neurosci.* **24**, 5445–5456 (2004).
 73. Ben-Arie, N. *et al.* Functional conservation of *atonal* and *Math1* in the CNS and PNS. *Development* **127**, 1039–1048 (2000).
 74. Wallis, D. *et al.* The zinc finger transcription factor Gfi1, implicated in lymphomagenesis, is required for inner ear hair cell differentiation and survival. *Development* **130**, 221–232 (2003).
 75. Cahusac, P. M. & Senok, S. S. Metabotropic glutamate receptor antagonists selectively enhance responses of slowly adapting type I mechanoreceptors. *Synapse* **59**, 235–242 (2006).
 76. Cahusac, P. M., Senok, S. S., Hitchcock, I. S., Genever, P. G. & Baumann, K. I. Are unconventional NMDA receptors involved in slowly adapting type I mechanoreceptor responses? *Neuroscience* **133**, 763–773 (2005).
 77. Fagan, B. M. & Cahusac, P. M. Evidence for glutamate receptor mediated transmission at mechanoreceptors in the skin. *Neuroreport* **12**, 341–347 (2001).
 78. Haeberle, H. *et al.* Molecular profiling reveals synaptic release machinery in Merkel cells. *Proc. Natl Acad. Sci. USA* **101**, 14503–14508 (2004).
 79. Tachibana, T. & Nawa, T. Immunohistochemical reactions of receptors to met-enkephalin, VIP, substance P, and CGRP located on Merkel cells in the rat sinus hair follicle. *Arch. Histol. Cytol.* **68**, 383–391 (2005).
 80. Jordt, S. E. & Julius, D. Molecular basis for species-specific sensitivity to 'hot' chili peppers. *Cell* **108**, 421–430 (2002).
 81. Bandell, M. *et al.* High-throughput random mutagenesis screen reveals TRPM8 residues specifically required for activation by menthol. *Nature Neurosci.* **9**, 493–500 (2006).
 82. Voets, T. *et al.* The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels. *Nature* **430**, 748–754 (2004).
 83. Feske, S. *et al.* A mutation in *Orai1* causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185 (2006).

Acknowledgements Research in the authors' laboratories is supported by the National Institutes of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to the authors (lumpkin@bcm.edu; caterina@jhmi.edu).

Pathogenesis and therapy of psoriasis

Michelle A. Lowes¹, Anne M. Bowcock² & James G. Krueger¹

Psoriasis is one of the most common human skin diseases and is considered to have key genetic underpinnings. It is characterized by excessive growth and aberrant differentiation of keratinocytes, but is fully reversible with appropriate therapy. The trigger of the keratinocyte response is thought to be activation of the cellular immune system, with T cells, dendritic cells and various immune-related cytokines and chemokines implicated in pathogenesis. The newest therapies for psoriasis target its immune components and may predict potential treatments for other inflammatory human diseases.

Psoriasis is a common skin disease that has been recognized since ancient times, when it was erroneously thought to be a variant of leprosy. Psoriasis affects about 25 million people in North America and Europe, and is probably the most prevalent immune-mediated skin disease in adults. It is an organ-specific autoimmune disease that is triggered by an activated cellular immune system^{1–6} and is similar to other immune-mediated diseases such as Crohn's disease, rheumatoid arthritis, multiple sclerosis and juvenile-onset diabetes. All of these fit the definition of an autoimmune disease as "a clinical syndrome caused by the activation of T cells and B cells, or both, in the absence of an ongoing infection or other discernable cause"⁷.

Because psoriasis occurs in an accessible organ it has been possible to study its cellular and genomic features in tremendous detail compared with those of other human autoimmune diseases. In turn, evolving pathogenic concepts are increasingly being tested directly in patients with psoriasis by administration of new therapies targeted to specific immune molecules. In this review we consider how interactions between resident skin cells and elements of the immune system — a complex network of cells and molecules that mediate innate and adaptive immunity — conspire to produce a disease that can last for decades in focal regions of the skin. We also briefly consider potential contributions of transmitted genes that increase susceptibility to psoriasis. More detailed discussion of the genetic defects and genomic pathways involved in psoriasis, and comparisons with other autoimmune diseases, are available elsewhere^{4,8}.

Clinical and histological appearance

Psoriasis vulgaris, the common form of psoriasis, is characterized by red, scaly, raised plaques. Although psoriasis vulgaris can occur in children, it often begins in late adolescence or early adulthood and then usually persists for life. Classic psoriasis vulgaris has a predilection for certain areas such as elbows, knees and the scalp. It may remain localized or become generalized over time. There are clinical variants of psoriasis, defined as subsets, with identical histopathological changes in the skin. Guttate psoriasis is characterized by small, scattered papules and is potentially linked to preceding streptococcal infections¹. Other recently described variants of psoriasis vulgaris include thick versus thin plaque disease⁹, and small versus large plaque disease¹⁰. A notable subset of patients with psoriasis develops psoriatic arthritis, a potentially debilitating illness².

Histologically, psoriasis has a defining appearance¹ (Fig. 1). There is marked thickening of the epidermis, due to increased proliferation of keratinocytes in the interfollicular epidermis, and epidermal rete — downward undulations of the epidermis — become very elongated

and form long, thin downward projections into the dermis (Fig. 1). The differentiation of keratinocytes is extensively altered in psoriasis⁴, paralleling 'regenerative maturation', an alternative cell differentiation programme that is transiently expressed during wound repair. Psoriatic plaques have surface scale, which is caused by aberrant terminal differentiation of keratinocytes. The granular layer of the epidermis, in which terminal differentiation begins, is greatly reduced or absent in psoriatic lesions. Consequently, a stratum corneum forms from incompletely differentiated keratinocytes that aberrantly retain a cell nucleus (this is known as parakeratosis, and the affected cells as parakeratotic keratinocytes). Scaling, and the consequential break in the protective barrier, are caused by failure of psoriatic corneocytes (terminally differentiated keratinocytes) to stack normally, secrete extracellular lipids and adhere to one another.

Other defining histological features of psoriasis include the presence of neutrophils within small foci in the stratum corneum and significant mononuclear infiltrates in the epidermis, which are detectable with immunostaining. In addition, there is marked infiltration of mononuclear leukocytes (T cells and dendritic cells, DCs) into the dermis and elongated/hyperplastic blood vessels in the papillary dermal region (between epidermal rete). Marked dilation of these vessels causes the visible redness of psoriatic skin lesions. Many lymphocytes, monocytes and neutrophils are clearly adherent to endothelial cells that acquire characteristics of high endothelial venules, which are usually found in lymph nodes. Endothelial cells are activated in psoriatic lesions, as is indicated by staining for intracellular adhesion molecule-1 (ICAM-1, also known as CD54), vascular cell adhesion molecule-1 (VCAM-1, or CD106) and E-selectin (CD62E)¹. Leukocytes can gain entry to skin parenchyma by transmigration through reactive vessels, but resident skin leukocytes might also expand to create the dense infiltrates seen in psoriatic lesions.

It is increasingly being recognized that even normal skin contains abundant stores of T lymphocytes¹¹ as well as resident populations of DCs¹², suggesting that skin might be a potential site for the direct triggering of recall immune responses. Experiments in which non-lesional skin from patients with psoriasis has been grafted to immunodeficient AGR mice have established the important principle that resident populations of T cells and DCs might be sufficient, when expanded, to induce psoriasis¹³. As illustrated in Fig. 1, psoriasis vulgaris lesions contain prominent aggregates of mononuclear leukocytes in the dermis that consist of hundreds to thousands of intermixed T cells and DCs, and these regions might function as organized lymphoid tissue that perpetuates immune infiltrates in psoriatic plaques¹⁴.

¹Laboratory for Investigative Dermatology, The Rockefeller University, 1230 York Avenue, Box 178, New York, New York 10021, USA. ²Department of Genetics, Washington University, School of Medicine, 4566 Scott Avenue, Saint Louis, Missouri 63110, USA.

The yin and yang of cellular interactions in psoriasis

Two fundamentally different cell types interact in the formation of a psoriatic lesion: epidermal keratinocytes and mononuclear leukocytes. Gene-expression programmes in these diverse cell types are likely to be influenced by distinct psoriasis susceptibility genes⁴. Whereas keratinocytes might be viewed only as bystander cells in terms of immune activation, it is more likely that they are active participants in the recruitment and activation of leukocytes in psoriatic lesions. Thus, there are two sets of interactive cellular responses in the psoriatic lesion that potentially create a yin/yang relationship — the balance between the activation of innate and acquired immune cell types, and the factors produced by epidermal keratinocytes that directly affect T cells and DCs, and vice versa (Fig. 2).

Effector cells of innate immunity in psoriatic lesions include neutrophils, plasmacytoid DCs and CD11c⁺ DCs. Because neutrophils are short-lived, they must be constantly recruited into lesions from blood stores. The chemokines interleukin-8 (IL-8) and growth-regulated oncogene- α (GRO- α , also known as CXCL1) — and possibly also S100A7/A8/A9 proteins — from keratinocytes produce a chemotactic gradient for the migration of neutrophils into the epidermis¹. BDCA-2⁺ CD123⁺ plasmacytoid DCs, which produce high levels of interferon- α (IFN- α) upon activation, have been proposed to have an important role in the

triggering of lesions¹⁵. CD11c⁺ DCs are increased in psoriatic lesions and constitute a cell group roughly equal to T cells in overall abundance¹⁶. Although Langerhans cells and dermal DCs have long been recognized as the main types of DC in skin¹⁷, it is now clear that psoriatic lesions contain additional types of DC. CD11c⁺ (myeloid) DCs correspond to interstitial DCs in other tissues and are the most abundant DC type in the dermis¹⁶. In addition, plasmacytoid DCs and several populations of activated DCs are present in psoriatic lesions (Fig. 2).

In psoriatic lesions, CD11c⁺ DCs express high levels of tumour necrosis factor (TNF) and the enzyme inducible nitric oxide synthase (iNOS), and have been proposed to be the human equivalent of TIP-DCs (TNF- and iNOS-producing DCs), cells that have effector functions in clearing some bacterial infections in mice^{16,18}. In addition, CD11c⁺ DCs probably produce the cytokines IL-23 and IL-20, which have the potential to activate T cells and keratinocytes, respectively^{19,20}. A fraction of CD11c⁺ DCs also bear 'maturation' markers, such as DC-LAMP or CD83 (ref. 16), and so could function as conventional DCs in terms of presenting antigens to T cells for the triggering of acquired immune responses. In fact, the juxtaposition of T cells and mature DCs in dermal aggregates (Fig. 1), as well as the expression of lymphoid-organizing chemokines such as CCL19, CCL21, CXCL12 and CCL18 (ref. 21; Fig. 2), may well promote T-cell activation *in situ*^{14,22,23}.

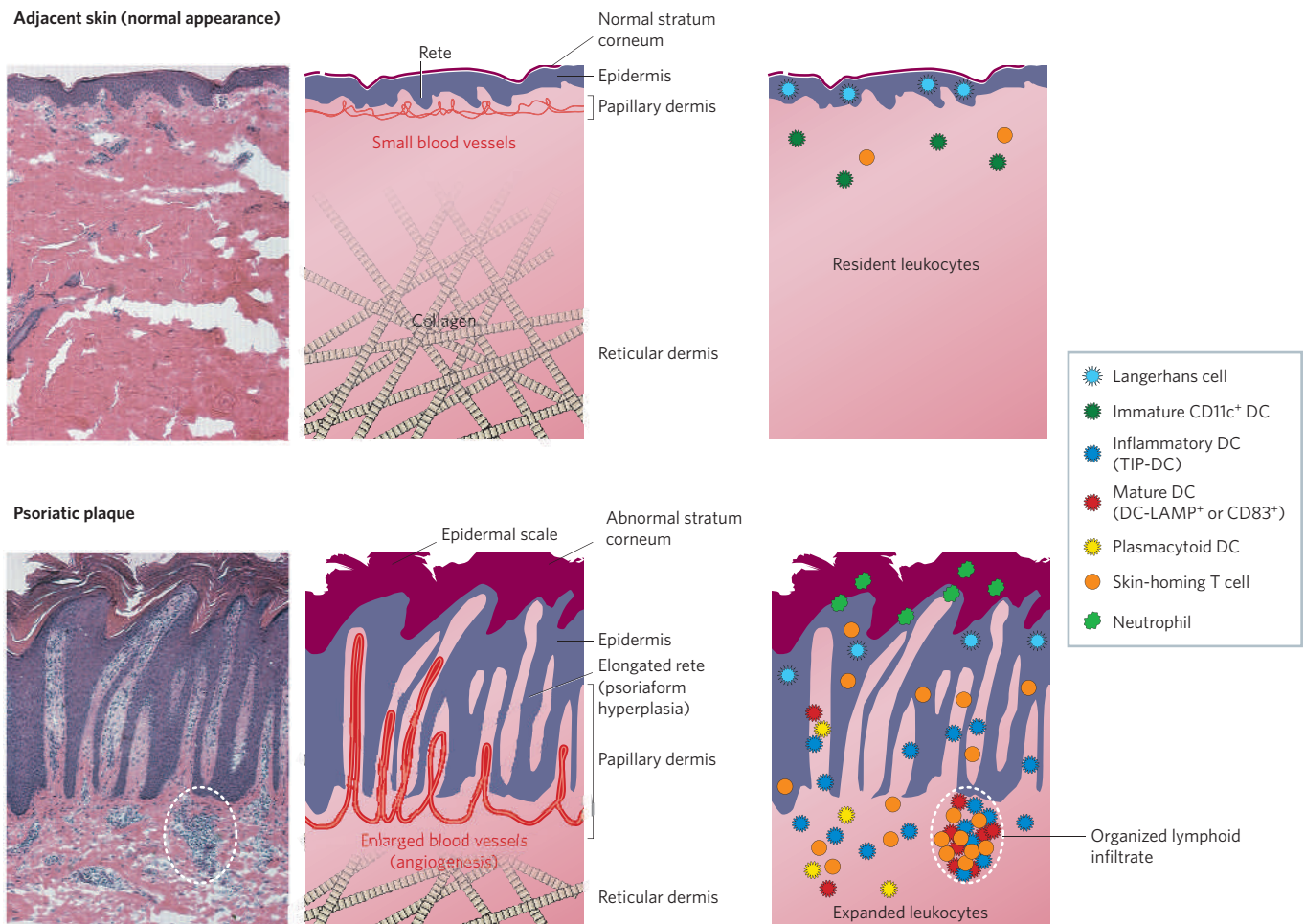


Figure 1 | Histological components of a mature psoriatic plaque compared with normal skin. Skin histology in normal skin and psoriatic lesions, with corresponding diagrams. The skin has three main layers. First there is the epidermis, which consists mainly of epithelial cells (keratinocytes). Second is the dermis, the bulk of which is made up of an extracellular matrix composed predominantly of collagens. This contains fibroblasts and a rich neurovascular network as well as many epidermal appendages that extend into the dermis, such as hair follicles, sebaceous glands and sweat glands. Third is the hypodermis, or layer of

subcutaneous adipose tissue with supporting stromal cells (not shown in figure). In psoriatic, cells of the stratum corneum (the outermost layer of the epidermis) stack abnormally, leading to the formation of scales, and the granular layer of the epidermis is much reduced. Epidermal rete are considerably elongated and blood vessels in the dermis are enlarged. Although normal skin contains notable numbers of resident and trafficking immune cells (and is an immune-competent organ), in psoriatic lesions the leukocyte number is significantly increased and many immune-related pathways are activated.

T cells in psoriatic lesions are polarized as T helper 1 (T_H1 ; $CD4^+$) and T cytotoxic (T_C1 ; $CD8^+$) subsets¹, but probably also include a separate population of T_H17 cells (induced by IL-23 in model systems)²⁴. Some $CD8^+$ T cells are specialized for homing into the epidermis through expression of $\alpha_E\beta_7$ integrin, which binds to E-cadherin on keratinocytes¹. In addition, many T cells express CD161 and other killer receptors, which might indicate a role for natural killer T cells in pathogenesis²⁵.

Keratinocyte products influence immune activation, and products of activated immunocytes alter keratinocyte responses, including the induction of new adhesion molecules for T cells (Fig. 2). Triggers for innate DCs might include heat-shock proteins or S100A12 produced by keratinocytes, or various environmental Toll-like receptor (TLR) agonists. Peptide antigens might also trigger conventional or acquired immune activation of T cells, as implied by the presence of clonal populations of T cells in lesions²⁶. Although antigen persistence could explain chronic immune activation, defective function of regulatory T (T_{reg}) cells has been suggested as another factor that might contribute to unbridled T-cell activation²⁷.

The sum of cellular interactions creates a tissue profile and a clinical phenotype, which we recognize as psoriasis vulgaris. Effective treatment with various immune antagonists breaks pathogenic immune activation and restores normal keratinocyte growth. In fact, the most important evidence that psoriasis is an immune-mediated disease comes from the finding that disease can be reversed with selective immune-targeted biological agents such as DAB389IL-2 (ref. 28) and CTLA4Ig (refs 29, 30).

Molecular pathways of inflammation

The ability to develop effective therapeutics by rational design is crucially dependent on elucidation of the molecular circuitry of inflammation in human autoimmune diseases. Cytokine interactions in psoriasis have previously been illustrated as a 'type-1 pathway', which assumes a linear relationship between proximal inducers (IL-23 or IL-12), production of IFN- γ and TNF by type-1 T cells, and downstream activation of numerous IFN-responsive genes through signal transducer and activator of transcription 1 (STAT1)¹⁴. Although this model is conceptually useful, it accounts for only a small fraction of the more than 1,300 genes that become upregulated in psoriatic lesions²¹. Figure 3 presents an alternative view of the inflammatory circuitry in psoriasis, which is more of a network or interactive model⁸. Clearly, STAT1, STAT3 and nuclear factor- κ B (NF- κ B) transcription factors are activated in psoriasis. Upstream activators may well be IFNs for STAT1, and TNF or IL-1 for NF- κ B, but more recently discovered cytokines such as IL-20 and IL-22 also have the ability to activate STAT and NF- κ B pathways^{20,31}, thus supporting the network concept.

Keratinocyte-derived cytokines such as platelet-derived growth factor (PDGF) and vascular endothelial growth factor (VEGF) influence the growth of supporting stromal cells. Activated stromal cells overproduce factors such as keratinocyte growth factor (KGF) that can induce proliferation of keratinocytes³². Many immune-derived cytokines, including IL-1, IL-6, IL-17, IL-19, IL-20, IL-22, TNF and IFNs, can also regulate keratinocyte proliferation, with some immune-derived cytokines clearly serving as alternative mitogens for this cell type. Antagonism of TNF

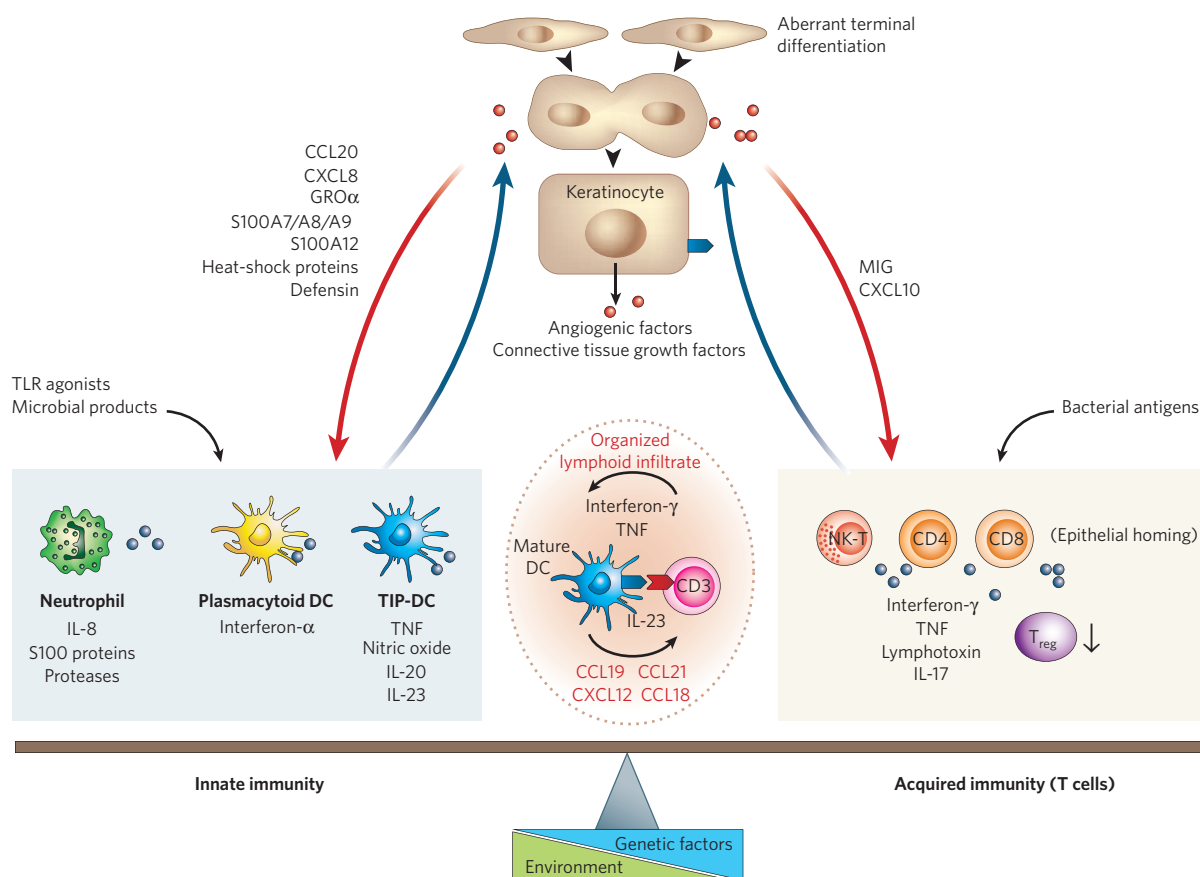


Figure 2 | A dynamic picture of the bidirectional flow of 'information' and cells in a mature psoriasis lesion: the yin and yang of psoriasis. There is close interdependence of the epidermis and dermal inflammatory infiltrate, as well as a balance between the innate and acquired immune systems. Chemokines produced by keratinocytes in the epidermis act on both the innate and acquired immune systems, stimulating DCs, neutrophils and other innate mediators as well as T cells. Keratinocytes also release cytokines and growth factors, leading to altered gene expression and

regenerative hyperplasia, and also to the induction of adhesion molecules for T cells on keratinocytes. Immune-system-derived cytokines, in turn, act on keratinocytes to either induce inflammatory genes or increase proliferation. Meanwhile, in the lymphoid-like tissue of the psoriatic dermis, molecules of the innate and acquired immune systems also interact. The genetic underpinnings of psoriasis are known to be complex, with ten or more susceptibility loci, and these probably interact with various environmental factors that act on the skin and/or immune system.

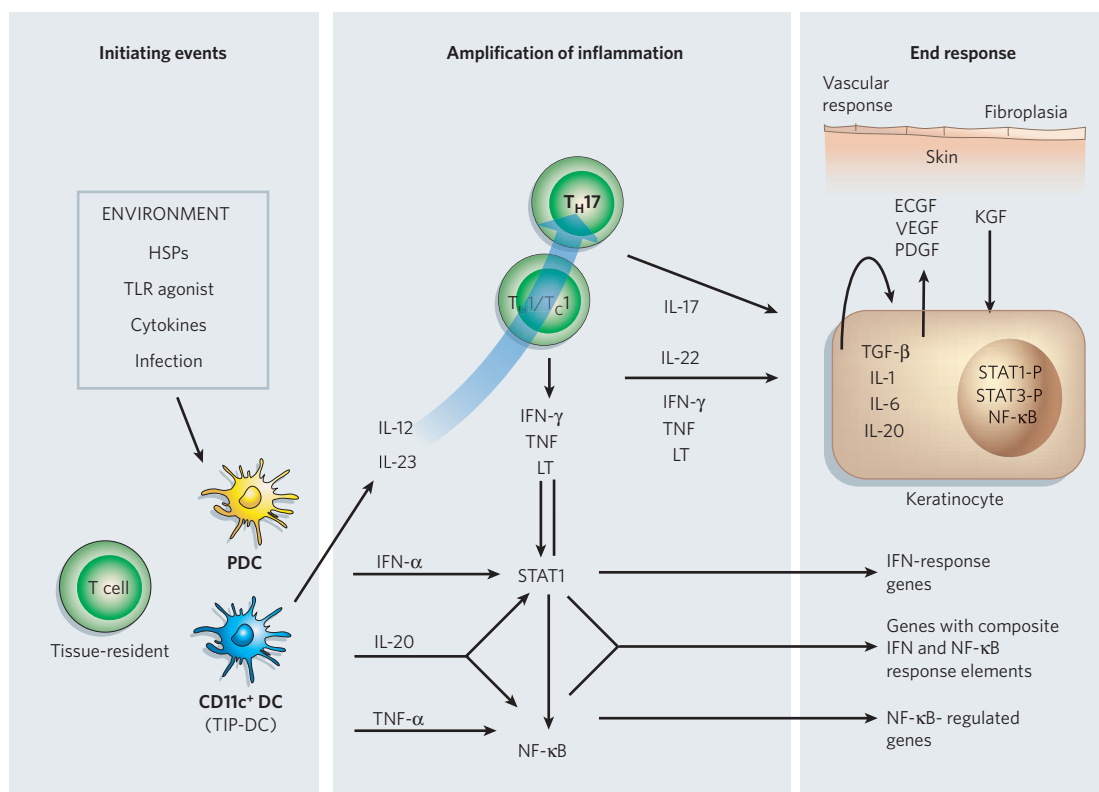


Figure 3 | Potential cytokine networks in psoriatic lesions. This figure shows some of the known interactions and products of cytokines and growth factors that are upregulated in psoriatic lesions. A notable fraction of the 1,300+ genes that are differentially expressed in psoriatic lesions are known to be regulated by STAT family or NF- κ B transcription factors, as well as by growth factors that are upregulated in psoriatic lesions. Key cytokines that can activate STAT or NF- κ B transcription factors (which are involved in amplifying inflammation) include TNFs, lymphotoxin (LT), IL-1, IL-17, IL-20, IL-22 and IFNs. Activated DCs may contribute IFN- α , IL-20, IL-12 and IL-23. T-cell activation through IL-12 or IL-23 (blue arrow)

leads to the synthesis of T-cell-derived inflammatory cytokines. Immune activation could be initiated by DC activation through pattern-recognition receptors, cytokines, or heat-shock proteins (HSPs), as well as by direct interaction with counter-receptors on T cells. Other cytokines synthesized by keratinocytes or stromal cells probably cross-regulate the epithelial-stromal (vascular) hyperplasia and fibroplasia that takes place in psoriatic lesions. TGF- β , IL-1, IL-6 and IL-20 may act as keratinocyte autocrine and/or paracrine growth factors. Certainly an equally complex set of chemokine interactions exists, as at least 15 chemokines have increased expression in psoriatic lesions²¹ and many other interactive pathways probably coexist.

and IL-12, and/or IL-23 (p40) cytokines with antibodies or fusion proteins, can break the activated pathways shown in Fig. 3, so the concept of proximal versus distal inflammatory regulators probably still holds even with more complex network circuitry.

Role of genetic factors in psoriasis

Psoriasis is essentially a disease of Caucasians, in whom its frequency is 1–2%. It is less common in Asians (about 0.1%) and is rarely seen in Africans³³. That psoriasis has a genetic basis has been accepted for many years⁴, and it is commonly thought of as a complex trait. So far, between 10 and 20 chromosome regions have been proposed to harbour psoriasis genes but less than a handful of genes have been identified^{4,8}. This is due, in part, to their low-risk effects and the limitations in the number of patients and families that have been studied.

One locus consistently identified in studies of psoriasis is the class I region of the major histocompatibility locus antigen cluster (MHC)⁴. However, its low penetrance — about 10% — indicates that other genetic and environmental factors are also involved³⁴. The identity of psoriasis susceptibility 1 (*PSORS1*) remains controversial. Although its association with human leukocyte antigen (HLA) Cw6 and psoriasis was reported more than 25 years ago³⁵, the extensive linkage disequilibrium across the class I region and its complex evolutionary history has made identification of the susceptibility variant(s) very difficult. Genes within this region lying about 160 kilobases telomeric to HLA-C, such as corneodesmosin (*CDSN*) and the α -helical coiled-coil rod (*HCR*), have been proposed as contenders⁴. A consensus is now beginning to emerge that supports the location of *PSORS1* as being closer to the region

harbouring HLA-C/HLA-B and excluding *CDSN* and *HCR*^{36–38}. However, whether *PSORS1* is a classical MHC allele, or a regulatory variant within this region, has not yet been agreed upon.

Other predisposing polygenes might affect the immune system or be involved in keratinocyte differentiation. Common variants in the *SLC9A3R1/NAT9* region and loss of a potential RUNX binding site have been described that could potentially affect regulation of the immune synapse^{4,39}. There has also been a report of an association of psoriasis with variant alleles of the lymphoid phosphatase *PTPN22* (ref. 40). This is also involved in regulation of the immune synapse and an R620W polymorphism is associated with at least four other autoimmune diseases³³. Associations with alleles encoding other components of the immune system such as IL-12 (ref. 41), IL-19/20 (ref. 42) and *IRF2* (ref. 43) have also been described.

Some genetic variants such as those from the epidermal differentiation complex (EDC) might directly affect keratinocyte proliferation or differentiation⁴. How subtle alterations in keratinocyte differentiation interact with alterations in the immune system to lead to the development of an inflammatory skin disease will be an important area of research as genetics progresses to global association scans, attempting to identify most of the common alleles.

The locus responsible for rare instances in which many members of a family are affected by psoriasis — which occurs when the disease segregates as a Mendelian trait — has been mapped to chromosome 17q25 (refs 44–46). These families are from the United States, Taiwan and Israel. Affected members of the Israeli family have autosomal-dominant seborrhoea-like dermatitis with psoriasiform elements that segregates with

a frameshift mutation of zinc finger protein 750 (ZNF750). This is normally expressed in keratinocytes but not in fibroblasts and is barely detectable in CD4 lymphocytes. Thus, in this case, the primary defect is in the keratinocyte rather than the immune system.

Therapeutic engineering and model systems

It is important to understand that human skin is a complex organ comprising many distinct tissues, and that its structure is significantly different from the skin of most lower species. Compared with fur-bearing animals, human skin has broad areas of epidermis situated between hair follicles, known as interfollicular epidermis (Fig. 1; Box 1). There are many different skin diseases that involve altered growth of epidermal keratinocytes and inflammation in the interfollicular epidermis, and psoriasis and atopic eczema are common examples. These disorders do not appreciably alter the growth of keratinocytes in the follicular epithelium or the growth of hair. Other diseases can alter the growth of follicular epithelium, sebaceous glands or hair (the pilosebaceous unit), and many such conditions are associated with immune infiltrates in or around follicular structures. Psoriasis does not exist as a spontaneous disease in the skin of lower animals⁴⁷, but some features of psoriasis have been induced in murine skin by genetic or immune manipulations. Even so, the structure of murine skin imposes serious limitations on resultant cellular alterations and, so far, psoriasis has not been faithfully reproduced by manipulation of native skin in any lower species (Box 1).

In the future, we need better model systems that will help us dissect the interactions of many complex molecular pathways or networks in the skin. We also need better systems with which to test possible therapeutic targets for psoriasis and other inflammatory diseases. Mice engineered with various transgenes to produce both epithelial hyperplasia and cutaneous inflammation might help with the first problem, but such models usually do not have 'regulated' or reversible phenotypes, as is the case with psoriatic lesions. From the therapeutic engineering perspective, it is important to model the extent to which pathological cell activation in psoriasis can be reversed by effective therapeutics (Table 1). Unlike many other autoimmune diseases, skin tissue is not irreversibly damaged by inflammation, so a complete reversal to normal skin structure is

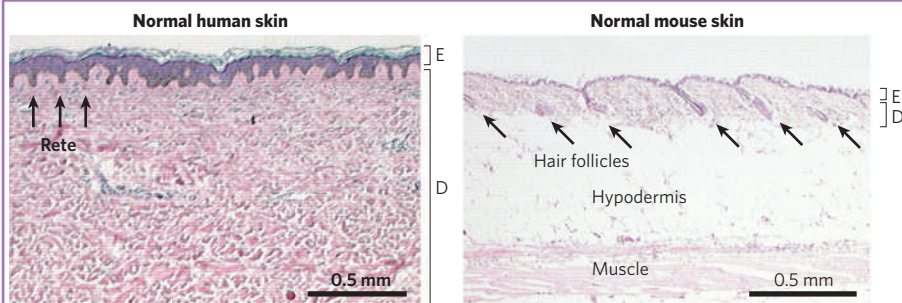
possible (Fig. 4). In this regard, models in which psoriatic skin has been xenotransplanted to immunodeficient mouse strains have produced clear examples of the prevention of disease induction by targeted therapeutics and of suppression of active psoriasis by antagonists of immune cells or inflammatory cytokines^{15,48}. More work might establish these models as predictive of clinical responses of psoriasis to various therapeutic agents, but at present the only reliable test of a new therapeutic is its ability to suppress psoriasis in a proof-of-concept clinical trial.

Therapeutics in psoriasis

Choice of treatment for psoriasis depends on many factors, including the extent of disease, its effect on a patient's life, and the patient's perception of their illness. For severe psoriasis, we now have biological therapies, which have been approved only during the past 3 years. Unlike earlier treatments for psoriasis, biological agents are proteins or antibodies that target specific molecules thought to be essential in psoriasis pathogenesis (Fig. 3). So far, these fall into two main groups, aimed either at specific inflammatory mediators such as TNF or more generally at T cells. The main concern about these and other biological agents is the effects of long-term chronic immunosuppression, which has the potential to increase infection and the risk of cancer. In clinical practice these drugs have been remarkably well tolerated, but we have only short-term safety data and need to continue to monitor these patients for long-term safety.

Table 1 lists selected systemic therapeutics used in psoriasis. Agents have been classified into four groups: first, agents used in early studies to establish the immunological basis of psoriasis (all of which are approved for other indications); second, common systemic immunosuppressives; third, new therapies that have recently been approved by regulatory agents in the United States or Europe; and, fourth, promising new therapies that are still under development. The main biological agents that are currently widely approved or in late-phase trials are discussed with respect to the relationship between therapeutics and pathogenesis. Such drugs include alefacept (LFA3-TIP, Amevive; Biogen), efalizumab (anti-CD11a, Raptiva; Genentec, Xoma, Serono), etanercept (Enbrel; Amgen, Wyeth), infliximab (Remicade; Centocor) and adalimumab (Humira; Abbott). See refs 49–51 for recent comprehensive reviews.

Box 1 | Key differences in the structure of human versus mouse skin



The images show representative micrographs of adult human and mouse skin photographed at the same magnification. Human skin contains sparse hair follicles that separate wide regions of interfollicular epidermis (no hair follicles are present in this image). By contrast, mouse skin has closely spaced hair follicles throughout (black arrows). Human skin has much thicker dermis (d) and epidermis (e), with many more cell layers in the epidermis. Note that the epidermal rete (downward projections of the epidermis, blue arrows), common in human epidermis, is absent in mouse skin. In addition, interfollicular epidermis in human skin has a distinct differentiation programme from keratinocytes in the follicular epithelium,

whereas this distinction is less clear in mouse epidermis.

A wide variety of transgenic and knockout mouse strains have been engineered in which growth factors, keratinocyte adhesion molecules, inflammatory mediators or leukocytes in the skin are altered^{47,62}. Often, perturbation of one of these pathways creates a phenotype with reactive keratinocyte hyperplasia, vascular proliferation and increased leukocyte infiltration of the skin. However, the dominant feature of almost all such models is hyperplasia of follicular keratinocytes. Tangential sections of this hyperplastic follicular epithelium can resemble elongated rete seen in psoriatic lesions (Fig. 1),

and be confused with the papillary elongation seen in psoriasis. In addition, many of these models have hair loss (alopecia) or even neonatal lethality⁶³, which are not normally associated with human psoriasis. Frequently, the 'psoriasis-like' phenotypes share features with other inflammatory skin disorders, such as atopic eczema, or inflammatory diseases of hair follicles. Thus, the pathology of models is often a unique phenotype that does not represent any human skin disease.

Occasionally, psoriasis-like phenotypes are created by molecular alterations that are uncommon in psoriatic plaques^{64,65}, and this leads to unnecessary confusion about the human condition and its underlying pathogenesis. In general, it is important to explore pathways consistently detected in psoriatic lesions. In addition, inflammatory skin models should be evaluated for reversal with common drugs used to treat psoriasis vulgaris; so far not many have been. Perhaps of most interest and therapeutic relevance are models in which human skin or psoriatic lesions are transplanted to immunodeficient mouse strains. In these cases, it has been possible to reproduce almost the full spectrum of cellular and histological changes that define psoriasis^{13,15}.

In a subset of patients with severe psoriasis, alefacept is a highly effective therapy and gives relatively long remissions in patients. Alefacept is a fusion protein that contains the extracellular domain of CD58 (LFA3) and binds to the surface co-stimulatory molecule CD2. The main cell types expressing CD2 are T cells and natural killer (NK) cells, but a small population of circulating CD14⁺ DCs are also CD2⁺. The main early hypotheses of alefacept's mechanism of action involved the bridging of T cells and NK cells by binding CD2 and the Fc receptor (FcR), respectively, leading to T-cell apoptosis⁵². This might explain the T-cell reductions associated with alefacept therapy, although we have not been able to demonstrate apoptosis in circulating cells (M.A.L. and J.G.K., unpublished observations). We have shown that in responding patients, alefacept induces a parallel reduction in CD11c⁺ and CD83⁺ DCs as well as T cells. A range of inflammatory genes such as IFN- γ , STAT1, MIG (CXCL9), iNOS, IL-8, and IL-23 subunits, as well as IL-20, are also reduced^{20,53}. It seems that T cells are the primary target for therapy, but that DCs and a spectrum of type-1 inflammatory genes are coordinately suppressed.

Efalizumab is an example of an agent that was also designed to interfere with T-cell adhesion and co-stimulation. It is effective in a subset of patients with severe psoriasis, although, like other therapies, it often requires long-term treatment to maintain disease control. Efalizumab is a humanized murine monoclonal antibody that targets CD11a, which forms a heterodimer with the β_2 integrin CD18 to form LFA-1. The CD11a/CD18 molecule is selectively expressed by T cells, and binds to ICAM-1 and 2. This interaction permits T-cell adhesion to ICAM⁺ DCs during the initial generation of immune responses in lymph nodes and is important in the skin during T-cell migration from the blood into dermis, local DC-activation of T cells, and T-cell entry into the epidermis. Administration of efalizumab induces a peripheral leukocytosis (predominantly of CD8⁺ memory cells⁵⁴), which is probably due to blockade

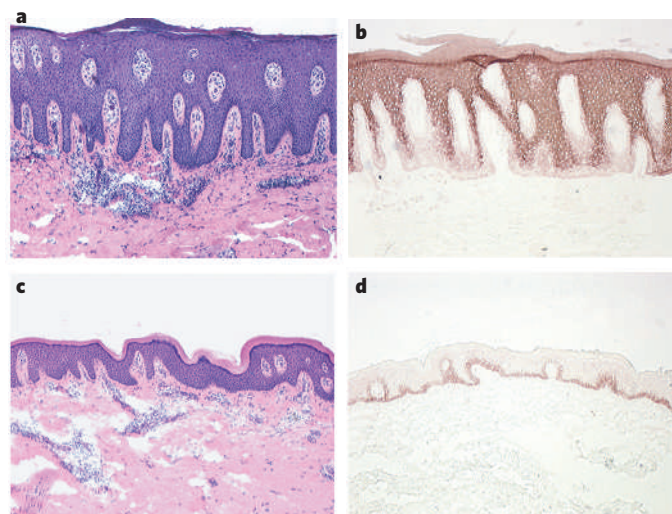


Figure 4 | Effective treatments are available for psoriasis and reverse the disease phenotype. **a, b**, Psoriatic lesions before treatment. **c, d**, Psoriatic lesions after 12 weeks of treatment with the T-cell-targeted monoclonal antibody efalizumab (anti-CD11a), showing reversibility and normalization of cutaneous histology. Panels **a** and **c** are stained with haematoxylin and eosin. Haematoxylin (purple) stains the chromatin of nuclei and eosin (pink) stains cytoplasmic material, connective tissue and collagen. Panels **b** and **d** are stained for keratin 16, a marker of epidermal regenerative maturation.

of the LFA-1/ICAM-1 interaction between T cells and endothelial cells. In addition, efalizumab therapy in patients with psoriasis causes a significant reduction in CD11c⁺ TIP-DCs¹⁶, and the 'switching off' of the mediators produced by this important cell type might also explain the clinical response.

The TNF inhibitors have greatly increased the treatment choices for patients with severe psoriasis. At present there are two FDA-approved agents for psoriasis (infliximab and etanercept) and one agent in late-phase trials (adalimumab). These TNF inhibitors have impressive disease control rates depending on the agent, formulation, dose and length of treatment. Etanercept is a human TNF receptor and immunoglobulin fusion protein that binds TNF and lymphotoxin- α and prevents their biological activity. Infliximab is a chimaeric human–murine monoclonal antibody against TNF that can bind both soluble and receptor-bound TNF. Adalimumab is the first fully human recombinant anti-TNF antibody, and, theoretically, has similar actions and effects to infliximab.

The use of targeted immune antagonists has tremendous potential not only for the treatment of psoriasis but also for the study of pathogenic contributions of specific immune molecules or pathways in autoimmune diseases such as psoriasis. The response of psoriasis to three distinct TNF inhibitors, which probably block the interaction of soluble TNF with TNF receptors on target cells, certainly suggests that this cytokine has a key role in disease pathogenesis. When coupled with cellular and genomic analyses of how inflammatory pathways collapse in response to targeted agents, we can learn more about how individual molecules influence the complex inflammatory networks that are apparent in psoriasis. For example, the progressive changes in inflammatory cytokines and chemokines induced by etanercept in psoriatic lesions suggests that TNF strongly regulates some proximal cytokines — for example, IL-1 and IL-8 — but has more complex long-range interactions to support inflammation driven by IFN- γ and STAT pathways, as well as chemokines that are thought to regulate T cells and DC interactions in the skin⁵⁵. In addition, several products that are inhibited by etanercept, such as iNOS and IL-23, are products of DCs that strongly express TNF (TIP-DCs)¹⁶ and are likely to be regulated by this cytokine.

However, the therapeutic actions of TNF inhibitors might not be as simple as blockade of the soluble cytokine. The antibody-based inhibitors (infliximab and adalimumab) have the potential to bind pro-TNF

Table 1 | Examples of systemic therapeutics for psoriasis vulgaris

Generic name (trade name)	Target	Status
Agents used in early studies to establish immunological basis of psoriasis		
Denileukin diftitox*/DAB389IL-2 (Ontak)	CD25 (toxic)	
Abatacept/CTLA4Ig (Orencia)	CD80 and CD86	
Tacrolimus/FK506 (Prograf)	Calcineurin	
Daclizumab*† (Zenepax)	CD25 (antagonist)	
Basiliximab*† (Simulect)	CD25 (antagonist)	
Widely used systemic agents approved for use (immunosuppressives)		
Cyclosporine* (Neoral, Gengraf)	Calcineurin	Widely used
Methotrexate (Rheumatrex, Trexall)	Leukocytes	Widely used
Fumarates*	T cells	Widely used in Germany
Approved biological agents		
Alefacept* (Amevive)	CD2	FDA approved
Efalizumab*† (Raptiva)	CD11a (LFA-1)	FDA and EMEA approved
Infliximab†‡ (Remicade)	TNF	FDA and EMEA approved
Etanercept‡ (Enbrel)	TNF, lymphotoxin	FDA and EMEA approved
Drugs/biological agents under investigation (human or murine trials)		
Adalimumab†‡ (Humira)	TNF	In clinical trials ⁵⁹ (FDA approved for psoriatic arthritis)
Pimecrolimus	Calcineurin	In clinical trials ⁶⁰
Cent-1275†‡	IL-12/23p40	In clinical trials ⁶¹
ABT-874†‡	IL-12/23p40	In clinical trials
146B7†‡	IL-15	In clinical trials ⁴⁸

*T-cell targeted.

†Monoclonal antibodies.

‡Cytokine inhibitors.

on the cell surface and TNF in its receptor-bound forms, and might therefore modify the biology of TNF⁺ cells through ligation of surface complexes or even induction of apoptosis⁵⁶. All of the TNF inhibitors have immunoglobulin domains that bind FcRs, which are expressed by several cell types but especially by DCs in psoriatic lesions. Some FcRs, particularly CD32b, have the potential to suppress immune responses when activated by antibody binding⁵⁷. Thus, suppression of immune circuits at the 'whole cell' level or through the elimination of TNF⁺ leukocytes could have the ability to suppress immune reactions that are not strictly TNF-dependent. The same arguments hold for antibody-like therapeutics to other targeted pathways. Therefore, we must couple the testing of targeted therapeutics with good cellular and molecular studies of skin lesions and circulating leukocytes to understand fully how inflammatory circuits are being affected by these agents. At the same time we need to gain a better understanding of how normal, protective immune responses might be affected during a lifetime of treatment with immune-regulating agents. Ideally, we might be able to find therapeutic agents that affect only pathological immune reactions and do not suppress protective cellular immunity. However, we still need a deeper understanding of upstream and downstream molecular interactions of inflammatory cytokines, chemokines and regulatory receptors in this disease.

Unresolved issues and questions

How do we move forward to gain a greater understanding of this unique disease of humans? First, it is likely that we have much to learn from careful mechanistic studies of existing targeted therapeutics in translational (clinical) studies. The introduction of new immune antagonists to additional targets will help to refine our understanding of the complex interactions that exist in psoriasis. Fortunately, an array of powerful molecular biological and genetic tools can be brought to the study of human tissue in an accessible organ within the context of therapeutic manipulation of disease activity.

Second, animal models of skin inflammation, even if they do not precisely reproduce psoriasis, have tremendous potential to help establish how complex inflammatory circuits are regulated. Animal models can also help us determine how psoriasis susceptibility genes might influence or dysregulate normal immune reactions or responses of skin-resident cells to immune triggers. In this case it is extremely helpful to model molecular or cellular changes that are consistently detected in psoriatic lesions. It would also be useful to have better information about how molecular alterations in murine model systems parallel the broad set of gene transcriptional alterations detected in psoriatic lesions through genome-level expression analyses. Unfortunately, none of the model systems have the level of genomic information that exists in psoriatic lesions studied from patients. In addition, it is not at all clear whether the repertoire of leukocyte subsets, including several distinct types of DC seen in psoriatic lesions, is represented in mouse skin⁵⁸.

Third, xenotransplantation of normal versus genetically affected skin of patients with psoriasis is a means of testing cellular or molecular interactions in ways that might not be possible in clinical studies, and this approach does have the advantage that the relevant human cell subsets are directly studied¹³. We are not yet at the point at which skin tissue can be fully reconstructed from cultured skin cells, but much progress is being made in this area and might be applied to pathogenic or therapeutic dissection in the future.

Hopefully, it can be appreciated that we have only begun to scratch the surface of an important skin disease. The progress we have made in understanding psoriasis leads only to a set of larger and more difficult questions for the future. The outstanding issues and questions might be broadly considered in five categories. First, what are the key triggers of the cellular inflammatory response going on in the skin and how do they interact with genetic susceptibility factors? Is this an autoimmune disease with self-reactivity to a conventional antigen, or a disease driven by endogenous or exogenous activators of innate immunity?

Second, why do wound healing responses of keratinocytes or immune-activation responses of leukocytes fail to terminate in psoriatic lesions,

as they would normally upon successful wound repair or elimination of an immune-activating pathogen? Or, to put it another way, is psoriasis a disease of too much immune stimulation or a problem in the response or downregulation of cell reactions to 'normal' stimulation? The potential for defective function of regulatory T cells, DCs or other cells that cross-regulate immune responses needs continued study.

Third, how do psoriasis susceptibility genes actually cause the broader set of cellular and transcriptional alterations that define psoriasis and make it different from other inflammatory diseases that also have genetic links, some of which overlap with psoriasis³³?

Fourth, how can we predict which patients might respond to expensive biological therapies? In psoriasis and other autoimmune diseases substantial genetic heterogeneity is apparent between patients. Although some genetic factors might encode components of the same biochemical pathway, and thus not require a different treatment regimen to halt disease, it is likely that genetic heterogeneity leads to subtle differences in disease pathogenesis, requiring different treatment regimens. As with other complex traits, the results of genome-wide methods, which will elucidate the genetic variations responsible for disease susceptibility and drug response, will allow personalized medicine to begin in earnest.

Finally, what are the differential effects of new targeted agents on pathological versus protective cellular immunity and, as a related issue, will we be able to safely alter the activity of the immune system over potentially decades of treatment? Whatever the ultimate answers to these questions hold, we are at an exciting time at which science and medicine converge to produce direct benefit to millions of affected individuals. ■

Note added in proof: A recent study⁶⁶ shows that IL-23 induces marked hyperplasia in epidermal keratinocytes in murine skin, and results suggest that this effect is mediated to a significant extent through IL-22 produced by T_H17 T cells. However, keratinocyte hyperplasia is still present on an *IL22*-null background, which suggests that IL-23 or other factors independent of IL-22 also stimulate keratinocyte proliferation.

1. Krueger, J. G. The immunologic basis for the treatment of psoriasis with new biologic agents. *J. Am. Acad. Dermatol.* **46**, 1–23 (2002).
2. Lebwohl, M. Psoriasis. *Lancet* **361**, 1197–1204 (2003).
3. Nicklöff, B. J. & Nestle, F. O. Recent insights into the immunopathogenesis of psoriasis provide new therapeutic opportunities. *J. Clin. Invest.* **113**, 1664–1675 (2004).
4. Bowcock, A. M. & Krueger, J. G. Getting under the skin: the immunogenetics of psoriasis. *Nature Rev. Immunol.* **5**, 699–711 (2005).
5. Schon, M. P. & Boehncke, W. H. Psoriasis. *N. Engl. J. Med.* **352**, 1899–1912 (2005).
6. Gaspari, A. A. Innate and adaptive immunity and the pathophysiology of psoriasis. *J. Am. Acad. Dermatol.* **54**, S67–S80 (2006).
7. Davidson, A. & Diamond, B. Autoimmune diseases. *N. Engl. J. Med.* **345**, 340–350 (2001).
8. Liu, Y., Krueger, J. G. & Bowcock, A. M. Psoriasis: genetic associations and immune system changes. *Genes Immunity* advance online publication (9 November 2006) doi:10.1038/sj.gene.6364351.
9. Christensen, T. E. *et al.* Observations of psoriasis in the absence of therapeutic intervention identifies two unappreciated morphologic variants, thin-plaque and thick-plaque psoriasis, and their associated phenotypes. *J. Invest. Dermatol.* **126**, 2397–2403 (2006).
10. Lew, W., Lee, E. & Krueger, J. G. Psoriasis genomics: analysis of proinflammatory (type 1) gene expression in large plaque (Western) and small plaque (Asian) psoriasis vulgaris. *Br. J. Dermatol.* **150**, 668–676 (2004).
11. Clark, R. A. *et al.* The vast majority of CLA⁺ T cells are resident in normal skin. *J. Immunol.* **176**, 4431–4439 (2006).
12. Boyman, O. *et al.* Activation of dendritic antigen-presenting cells expressing common heat shock protein receptor CD91 during induction of psoriasis. *Br. J. Dermatol.* **152**, 1211–1218 (2005).
13. Boyman, O. *et al.* Spontaneous development of psoriasis in a new animal model shows an essential role for resident T cells and tumor necrosis factor- α . *J. Exp. Med.* **199**, 731–736 (2004).
14. Lew, W., Bowcock, A. M. & Krueger, J. G. Psoriasis vulgaris: cutaneous lymphoid tissue supports T-cell activation and 'Type 1' inflammatory gene expression. *Trends Immunol.* **25**, 295–305 (2004).
15. Nestle, F. O. *et al.* Plasmacytoid dendritic cells initiate psoriasis through interferon- α production. *J. Exp. Med.* **202**, 135–143 (2005).
16. Lowes, M. A. *et al.* Increase in TNF- α and inducible nitric oxide synthase-expressing dendritic cells in psoriasis and reduction with efalizumab (anti-CD11a). *Proc. Natl Acad. Sci. USA* **102**, 19057–19062 (2005).
17. Larregina, A. T. & Faló, L. D. Changing paradigms in cutaneous immunology: adapting with dendritic cells. *J. Invest. Dermatol.* **124**, 1–12 (2005).
18. Serbina, N. V., Salazar-Mather, T. P., Biron, C. A., Kuziel, W. A. & Pamer, E. G. TNF/ α /iNOS-producing dendritic cells mediate innate immune defense against bacterial infection. *Immunity* **19**, 59–70 (2003).
19. Lee, E. *et al.* Increased expression of interleukin 23 p19 and p40 in lesional skin of patients with psoriasis vulgaris. *J. Exp. Med.* **199**, 125–130 (2004).

20. Wang, F. *et al.* Prominent production of IL-20 by CD68⁺/CD11c⁺ myeloid-derived cells in psoriasis: gene regulation and cellular effects. *J. Invest. Dermatol.* **126**, 1590–1599 (2006).
21. Zhou, X. *et al.* Novel mechanisms of T-cell and dendritic cell activation revealed by profiling of psoriasis on the 63,100-element oligonucleotide array. *Physiol. Genomics* **13**, 69–78 (2003).
22. Weninger, W. *et al.* Naive T cell recruitment to nonlymphoid tissues: a role for endothelium-expressed CC chemokine ligand 21 in autoimmune disease and lymphoid neogenesis. *J. Immunol.* **170**, 4638–4648 (2003).
23. Weninger, W. & von Andrian, U. H. Chemokine regulation of naive T cell traffic in health and disease. *Semin. Immunol.* **15**, 257–270 (2003).
24. McKenzie, B. S., Kastelein, R. A. & Cua, D. J. Understanding the IL-23–IL-17 immune pathway. *Trends Immunol.* **27**, 17–23 (2006).
25. Nickoloff, B. J., Bonish, B., Huang, B. B. & Porcelli, S. A. Characterization of a T cell line bearing natural killer receptors and capable of creating psoriasis in a SCID mouse model system. *J. Dermatol. Sci.* **24**, 212–225 (2000).
26. Prinz, J. C. *et al.* T cell clones from psoriasis skin lesions can promote keratinocyte proliferation *in vitro* via secreted products. *Eur. J. Immunol.* **24**, 593–598 (1994).
27. Sugiyama, H. *et al.* Dysfunctional blood and target tissue CD4⁺CD25^{high} regulatory T cells in psoriasis: mechanism underlying unrestrained pathogenic effector T cell proliferation. *J. Immunol.* **174**, 164–173 (2005).
28. Gottlieb, S. L. *et al.* Response of psoriasis to a lymphocyte-selective toxin (DAB389IL-2) suggests a primary immune, but not keratinocyte, pathogenic basis. *Nature Med.* **1**, 442–447 (1995).
29. Abrams, J. R. *et al.* CTLA4Ig-mediated blockade of T-cell costimulation in patients with psoriasis vulgaris. *J. Clin. Invest.* **103**, 1243–1252 (1999).
30. Abrams, J. R. *et al.* Blockade of T lymphocyte costimulation with cytotoxic T lymphocyte-associated antigen 4-immunoglobulin (CTLA4Ig) reverses the cellular pathology of psoriatic plaques, including the activation of keratinocytes, dendritic cells, and endothelial cells. *J. Exp. Med.* **192**, 681–694 (2000).
31. Wolk, K. *et al.* IL-22 increases the innate immunity of tissues. *Immunity* **21**, 241–254 (2004).
32. Finch, P. W., Murphy, F., Cardinale, I. & Krueger, J. G. Altered expression of keratinocyte growth factor and its receptor in psoriasis. *Am. J. Pathol.* **151**, 1619–1628 (1997).
33. Bowcock, A. M. The genetics of psoriasis and autoimmunity. *Annu. Rev. Genomics Hum. Genet.* **6**, 93–122 (2005).
34. Elder, J. T. *et al.* The genetics of psoriasis. *Arch. Dermatol.* **130**, 216–224 (1994).
35. Tiilikainen, A., Lassus, A., Karvonen, J., Vartiainen, P. & Julin, M. Psoriasis and HLA-Cw6. *Br. J. Dermatol.* **102**, 179–184 (1980).
36. Veal, C. D. *et al.* Family-based analysis using a dense single-nucleotide polymorphism-based map defines genetic variation at PSORS1, the major psoriasis-susceptibility locus. *Am. J. Hum. Genet.* **71**, 554–564 (2002).
37. Nair, R. P. *et al.* Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am. J. Hum. Genet.* **78**, 827–851 (2006).
38. Helms, C. *et al.* Localization of PSORS1 to a haplotype block harboring HLA-C and distinct from corneodesmosin and HCR. *Hum. Genet.* **118**, 466–476 (2005).
39. Helms, C. *et al.* A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. *Nature Genet.* **35**, 349–356 (2003).
40. Huffmeier, U. *et al.* Evidence for susceptibility determinant(s) to psoriasis vulgaris in or near PTPN22 in German patients. *J. Med. Genet.* **43**, 517–522 (2006).
41. Tsunemi, Y. *et al.* Interleukin-12 p40 gene (IL12B) 3′-untranslated region polymorphism is associated with susceptibility to atopic dermatitis and psoriasis vulgaris. *J. Dermatol. Sci.* **30**, 161–166 (2002).
42. Koks, S. *et al.* Combined haplotype analysis of the interleukin-19 and -20 genes: relationship to plaque-type psoriasis. *Genes Immun.* **5**, 662–667 (2004).
43. Foerster, J. *et al.* Evaluation of the IRF-2 gene as a candidate for PSORS3. *J. Invest. Dermatol.* **122**, 61–64 (2004).
44. Tomfohrde, J. *et al.* Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q. *Science* **264**, 1141–1145 (1994).
45. Hwu, W. L. *et al.* Mapping of psoriasis to 17q terminus. *J. Med. Genet.* **42**, 152–158 (2005).
46. Birnbaum, R. Y. *et al.* Seborrhea-like dermatitis with psoriasiform elements caused by a mutation in ZNF750, encoding a putative C2H2 zinc finger protein. *Nature Genet.* **38**, 749–751 (2006).
47. Nestle, F. O. & Nickoloff, B. J. From classical mouse models of psoriasis to a spontaneous xenograft model featuring use of AGR mice. *Ernst Schering Res. Found. Workshop* 203–212 (2005).
48. Villadsen, L. S. *et al.* Resolution of psoriasis upon blockade of IL-15 biological activity in a xenograft mouse model. *J. Clin. Invest.* **112**, 1571–1580 (2003).
49. Weinberg, J. M., Bottino, C. J., Lindholm, J. & Buchholz, R. Biologic therapy for psoriasis: an update on the tumor necrosis factor inhibitors infliximab, etanercept, and adalimumab, and the T-cell-targeted therapies efalizumab and alefacept. *J. Drugs Dermatol.* **4**, 544–555 (2005).
50. Gottlieb, A. B. Psoriasis: emerging therapeutic strategies. *Nature Rev. Drug Discov.* **4**, 19–34 (2005).
51. Papp, K. A. The long-term efficacy and safety of new biological therapies for psoriasis. *Arch. Dermatol. Res.* **298**, 7–15 (2006).
52. Ellis, C. N. & Krueger, G. G. Treatment of chronic plaque psoriasis by selective targeting of memory effector T lymphocytes. *N. Engl. J. Med.* **345**, 248–255 (2001).
53. Chamian, F. *et al.* Alefacept reduces infiltrating T cells, activated dendritic cells, and inflammatory genes in psoriasis vulgaris. *Proc. Natl Acad. Sci. USA* **102**, 2075–2080 (2005).
54. Vugmeyster, Y. *et al.* Efalizumab (anti-CD11a)-induced increase in leukocyte numbers in psoriasis patients is preferentially mediated by blocked entry of memory CD8⁺ T cells into the skin. *Clin. Immunol.* **113**, 38–46 (2004).
55. Gottlieb, A. B. *et al.* TNF inhibition rapidly down-regulates multiple proinflammatory pathways in psoriasis plaques. *J. Immunol.* **175**, 2721–2729 (2005).
56. Kruger-Krasagakis, S., Galanopoulos, V. K., Giannikaki, L., Stefanidou, M. & Tosca, A. D. Programmed cell death of keratinocytes in infliximab-treated plaque-type psoriasis. *Br. J. Dermatol.* **154**, 460–466 (2006).
57. Boruchov, A. M. *et al.* Activating and inhibitory IgG Fc receptors on human DCs mediate opposing functions. *J. Clin. Invest.* **115**, 2914–2923 (2005).
58. Dupasquier, M., Stoltzner, P., van Oudenaren, A., Romani, N. & Leenen, P. J. Macrophages and dendritic cells constitute a major subpopulation of cells in the mouse dermis. *J. Invest. Dermatol.* **123**, 876–879 (2004).
59. Chen, D. M., Gordon, K., Leonardi, C. & Menter, M. A. Adalimumab efficacy and safety in patients with moderate to severe chronic plaque psoriasis: preliminary findings from a 12-week dose-ranging trial. *J. Am. Acad. Dermatol.* **50**, 1 (2004).
60. Gottlieb, A. B. *et al.* Oral pimecrolimus in the treatment of moderate to severe chronic plaque-type psoriasis: a double-blind, multicentre, randomized, dose-finding trial. *Br. J. Dermatol.* **152**, 1219–1227 (2005).
61. Toichi, E. *et al.* An anti-IL-12p40 antibody down-regulates type 1 cytokines, chemokines, and IL-12/IL-23 in psoriasis. *J. Immunol.* **177**, 4917–4926 (2006).
62. Xia, Y. P. *et al.* Transgenic delivery of VEGF to mouse skin leads to an inflammatory condition resembling human psoriasis. *Blood* **102**, 161–168 (2003).
63. Blumberg, H. *et al.* Interleukin 20: discovery, receptor identification, and role in epidermal function. *Cell* **104**, 9–19 (2001).
64. Zenz, R. *et al.* Psoriasis-like skin disease and arthritis caused by inducible epidermal deletion of Jun proteins. *Nature* **437**, 369–375 (2005).
65. Haider, A. S., Duculan, J., Whynot, J. A. & Krueger, J. G. Increased JunB mRNA and protein expression in psoriasis vulgaris lesions. *J. Invest. Dermatol.* **126**, 912–914 (2006).
66. Zheng, Y. *et al.* Interleukin-22, a T_H17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. *Nature advance online publication* (24 December 2006) doi:10.1038/nature05505.

Acknowledgements The authors and their primary research have been supported by grants from the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to J.G.K. (jgk@rockefeller.edu).

Progress and opportunities for tissue-engineered skin

Sheila MacNeil¹

Tissue-engineered skin is now a reality. For patients with extensive full-thickness burns, laboratory expansion of skin cells to achieve barrier function can make the difference between life and death, and it was this acute need that drove the initiation of tissue engineering in the 1980s. A much larger group of patients have ulcers resistant to conventional healing, and treatments using cultured skin cells have been devised to restart the wound-healing process. In the laboratory, the use of tissue-engineered skin provides insight into the behaviour of skin cells in healthy skin and in diseases such as vitiligo, melanoma, psoriasis and blistering disorders.

The overriding function of tissue-engineered skin is to restore barrier function to patients in whom this has been severely compromised. Figure 1 shows the structure of normal skin, and Fig. 2 shows an example of a patient in whom tissue-engineered skin has been used clinically. Any loss of full-thickness skin of more than 4 cm in diameter will not heal well without a graft¹. In cases in which considerable amounts of skin are needed, the 'gold standard' approach is to take split-thickness grafts that contain all of the epidermis but only part of the dermis. These are removed from healthy areas of the body and used to treat damaged areas. Patients will regrow an epidermis from the source sites if there are sufficient epidermal cells remaining in the residual dermis. Before tissue-engineered skin was available, surgeons had to avoid making the patient's condition acutely worse by removing too much epidermis from elsewhere on the body (burn fatalities after initial stabilization are usually due to an insufficient barrier failing to prevent the onset of bacterial sepsis).

Acute burns are not numerically significant in the western world^{2,3}, where they have proved to be largely avoidable by education and lifestyle, but they remain a major healthcare problem for developing countries. In these communities there are neither the economic resources nor the infrastructure to establish tissue-engineering technologies, but simple knowledge transfer could make a considerable difference. In the developed world, life expectancy and affluence have increased so markedly that chronic wounds associated with ageing and diabetes have become very significant⁴. These are expensive, both economically for the healthcare system and in human terms for the patient⁵. Repeated applications of skin cells, whether these be keratinocytes or fibroblasts, autologous (the patient's own) or allogeneic (from human donors), can all offer some benefit to chronic non-healing wounds in prompting them to restart healing. Here cultured cells are being used as biological 'factories' to assist the body's own wound repair mechanisms. Other areas of increasing clinical potential are applications in reconstructive surgery, scar revision, scar prevention, correction of pigmentation defects and the treatment of some blistering diseases (such as congenital naevi⁶⁻⁸).

This article highlights progress in this field, reviewing the underpinning cell biology, the key clinical landmarks and facilitating biomaterials, and examines the role of tissue-engineered skin in developing our understanding of skin biology.

Principles of skin tissue engineering

Skin comprises several different cell types. Keratinocytes are the most common cell type in the epidermis and form the surface barrier layer. Melanocytes are found in the lower layer of the epidermis and provide skin colour. Fibroblasts form the lower dermal layer and provide strength and resilience (Fig. 1). Most tissue-engineered skin is created by expanding skin cells in the laboratory (at a rate much greater than would be achieved on the patient) and using them to restore barrier function (the primary objective for burns patients) or to initiate wound healing (for chronic non-healing ulcers). Other uses include accelerating healing, reducing pain in superficial burns and correcting conditions in which healing has been suboptimal (for example, in scars, contractures or pigmentary defects). Skin needs to be capable of regeneration, so although synthetic materials can be used temporarily to provide a barrier, a dermal matrix or a transfer mechanism, for longer-term healing all synthetic materials must eventually be discarded or replaced by live skin cells.

Three factors are paramount in the development of tissue-engineered materials: the safety of the patient, clinical efficacy and convenience of use. Any cultured cell material carries the risk of transmitting viral or bacterial infection, and some support materials (such as bovine collagen and murine feeder cells) may also have a disease risk. There must be clear evidence that tissue-engineered materials provide benefit to the patient. Essential characteristics are that it heals well and has the physical properties of normal skin. To achieve effective healing the tissue-engineered products must attach well to the wound bed, be supported by new vasculature, not be rejected by the immune system and be capable of self repair throughout a patient's life. Finally, materials need to be convenient to use or they will not achieve clinical uptake.

Key developments

Keratinocytes were first reliably cultured in the laboratory about 30 years ago^{9,10}. This developed into the production of small sheets of cells two or three layers thick¹¹ (known as cultured epithelial autografts, CEAs) that were used to treat burns victims^{12,13}. The Timeline shows key events in the development of tissue-engineered skin. Evidence suggests that cultured skin cells returned to a patient retain the capacity to self-renew for a lifetime. Patients who received skin in the mid-1980s¹³ are alive today and have not required further treatment. Research

¹The Tissue Engineering Group, Department of Engineering Materials and Division of Biomedical Sciences and Medicine, Kroto Research Institute, North Campus, University of Sheffield, Broad Lane, Sheffield S3 7HQ, UK.

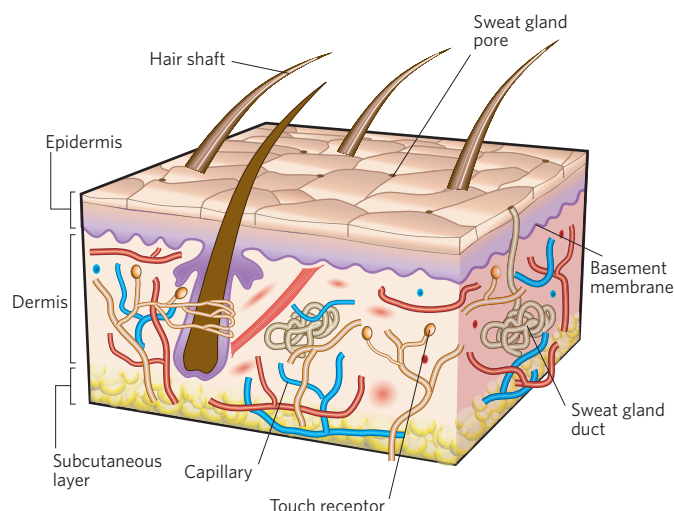


Figure 1 | The structure of human skin. This diagram of human skin shows the two main layers of skin — the upper epidermal barrier layer and the lower, much thicker, dermis. The epidermal barrier layer is relatively thin (0.1–0.2 mm in depth) and securely attached to the underlying dermis by a specialized basement membrane zone. This consists of several different types of collagen fibre, which attach cells securely to the underlying dermis, and is visible at the electron-microscope level. The dermis varies in thickness depending on its site in the body and is composed primarily of collagen I, with dermal inclusions of hair shafts and sweat glands, which are lined with epidermal keratinocytes. The dermis is well vascularized and also contains receptors for touch, temperature and pain. Keratinocytes in the epidermis rely solely on diffusion from the adjacent dermal capillary network. These cells progressively differentiate from cells in the basal layer, which is located on the basement membrane and gives rise to daughter keratinocytes, which are pushed upwards. These stratify, lose their nuclei and eventually become an integrated sheet of keratin, which is later shed. The upper keratinized epidermal layers provide the barrier layer, which resists bacterial entry and prevents fluid and electrolyte loss. (Image adapted, with permission, from ref. 79.)

in vitro has indicated that skin is the most accessible of all adult stem-cell populations^{14–16} (see page 834).

Without a well-vascularized dermal wound bed, cultured keratinocytes alone are of limited value in treating full-thickness burns¹⁷. In such cases it is necessary to replace both epidermal and dermal layers of skin. Accepted practice¹⁸ in these cases is to debride the burned tissue (often to the muscle fascia), provide a dermal equivalent^{17,19} and, once this has become well vascularized, provide an epidermal layer. This need to provide a dermis prompted the evaluation of a range of biomaterials, both natural and synthetic. Donor (cadaveric) skin is commonly used for managing serious burns and treating chronic wounds. It can be grafted onto full-thickness wounds and will become vascularized. The upper epidermal layer (which is not immune-system compatible) is removed after about 3 weeks, leaving behind a vascularized donor dermis, which provides an ideal wound bed for subsequent skin grafts or tissue-engineered skin graft substitutes¹⁷. Donor skin is currently an underused resource that could be especially valuable in the developing world. Skin banks can be established at much lower cost than that of purchasing other dermal substitutes. With good banking practices and sterilization²⁰ the risks of using donor skin can be reduced to that of using blood.

Another useful dermal substitute is Integra²¹. This was designed for the management of major burns and comprises bovine collagen and shark chondroitin sulphate with a silicone membrane that functions as a temporary barrier. The material is grafted onto the wound bed and under this membrane vasculogenesis occurs. After vascularization, the silicone barrier membrane is removed and replaced with a barrier of the patient's own cells — either a split-thickness skin graft or tissue-engineered material. Direct application of cultured cells to an Integra wound bed was

found to be problematic, because cells failed to adhere securely²². This led to the development of reconstructed tissue-engineered skin based on autologous keratinocytes, fibroblasts and bovine collagen sheets (refs 22–24; see ref. 25 for a review). These attached well to an Integra-treated wound bed. As our knowledge of burns treatments developed, the use of donor cells to treat chronic wounds was explored and there is now a range of products to reinitiate wound healing in such cases.

Current progress and challenges

Currently available tissue-engineered products can be categorized into three groups (Box 1). There are those that replace the epidermal layer only, those that provide a dermal substitute, and a small number that provide both. In some clinical conditions (such as non-healing ulcers and superficial burns) simply transferring laboratory-expanded cells can benefit patients, but the treatment of major full-thickness burns requires the replacement of both dermis and epidermis. There are four major challenges in this field: improving safety, finding a substitute for split-thickness grafts, improving angiogenesis in replacement tissue once it has been grafted to the wound bed, and improving ease of use.

Safety

The predominant culture methodology^{9,10} uses murine fibroblasts and bovine serum. This process pre-dated knowledge of prion diseases and concerns that murine cells might contain viruses able to transform human cells. To reduce risk, murine feeder fibroblasts are subjected to lethal gamma radiation and serum is sourced from countries where herds are free from bovine spongiform encephalitis (BSE). However, if BSE spreads worldwide, serum-free media will become a necessity. Early versions of this proved unreliable but more promising products, such as Epilife (for an example of use see ref. 26), now exist. As these are proprietary products, transfer to the clinic is restricted, but registration by the US Food and Drug Administration (FDA) is underway. There are also specific safety issues with natural biomaterials used in skin reconstruction. Xenotransplantation — the use of porcine tissue in humans — carries the risk of stimulating antibodies to the Gal epitope. This membrane glycoprotein is present in small amounts on pig tissue and has the potential to activate human complement, causing acute rejection²⁷, so particular care must be taken to remove this before use.

Donor cells need to be from screened donors to avoid viral diseases, and the products that use them (such as Dermagraft²⁸, Transcyte²⁹, Apligraf³⁰ and OrCel³¹) need to be from screened cell banks. Donor cells can benefit patients in two ways — they can act as biological factories to stimulate wound healing through the production of growth factors, and they can also be used as temporary dressings. Donor fibroblasts can be used both for chronic wound stimulation and in dermal replacement

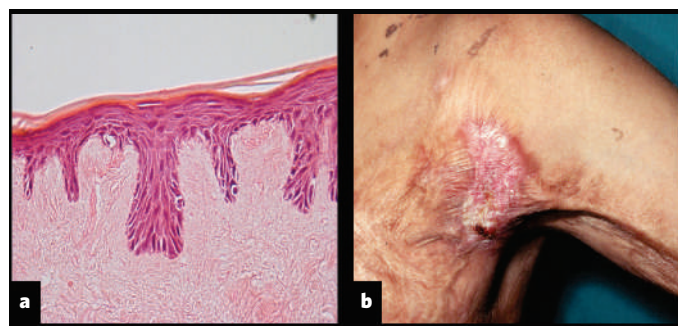
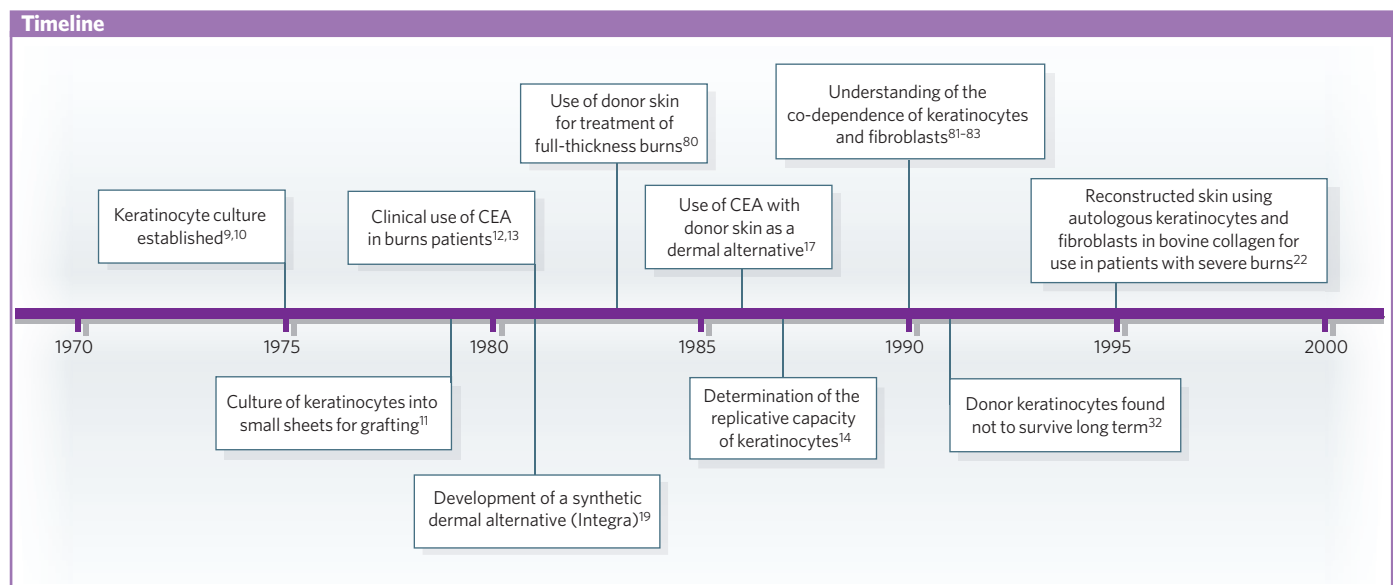


Figure 2 | An example of the clinical use of tissue-engineered skin.

a, The appearance of reconstructed tissue-engineered human skin based on sterilized de-epidermized acellular donor dermis to which the patient's own laboratory-expanded keratinocytes and fibroblasts were added. **b**, The appearance of this reconstructed skin 2 months after its use in elective surgery to correct contraction in a patient who had suffered severe skin contractions as a result of earlier burns injuries. The tissue-engineered skin graft was placed in the axilla to allow the patient a more normal range of movement. (Images reproduced, with permission, from ref. 39.)



for severe burns. However, there is no evidence that donor keratinocytes have ever been used successfully in the long-term treatment of full-thickness burns^{32,33}. One unanswered question in this area is whether donor fibroblasts survive long term when used clinically. Donor skin has been in use since the early 1980s^{17,34,35} and no problems have been documented in association with residual donor fibroblasts. Pig skin is also used as a temporary substitute (for example, Permacol³⁶). However, the development of a successful synthetic dermal substitute would considerably reduce many of these risks.

Split-thickness graft substitutes

Although split-thickness grafts remain the gold-standard treatment for many clinical conditions, they are not always available in sufficient quantity, giving rise to the clinical need for tissue-engineered alternatives. Commercial products are beginning to emerge to meet this need (such as Permaderm, previously known as Cincinnati skin substitute³⁷) but there are still considerable challenges to be met. When delivering keratinocytes to deep burns, the stability of the attachment of the keratinocytes to the underlying dermis is the key issue. It is very difficult for these cells to attach well (equivalent to early-stage basement membrane formation) unless the wound bed has been well prepared to achieve a vascularized dermis or dermal equivalent. A better approach is to synthesize a reconstructed skin in the laboratory in which attachment between keratinocytes and dermis can be secured^{22,38,39} (Fig. 2). Angiogenesis can also be a problem. The maximum thickness of skin-replacement material that can easily become vascularized is about 0.4 mm, and failure to take this into consideration when using reconstructed skin based on human allodermis can result in delayed angiogenesis and loss of grafts³⁹. In grafts thicker than 0.4 mm, new blood vessels cannot penetrate quickly enough to feed the epidermal layer. Reconstructed skin placed over fat or poorly vascularized wound beds will be lost. The Boyce group found that the most reliable 'take' of skin equivalents is obtained if the wound bed is pretreated with a product such as Integra²¹ to achieve good dermal vascularization before grafting²². Even after the successful restoration of barrier function (the immediate objective in treating burn injuries), aesthetic appearance can be far from ideal. For example, contracture of grafts³⁹ and abnormal pigmentation^{40,41} can occur. Clinicians and research scientists might consider wound healing and restoration of barrier function with full mobility to be great achievements, and contracture and abnormal pigmentation to be lesser considerations. However, for the patient these issues are important and more must be done to improve the quality of the final result.

Any three-dimensional (3D) reconstructed skin needs a dermal scaffold. The development of biomaterials for clinical use is an extensive

and rapidly expanding field (see ref. 42 for a review). In skin tissue engineering the ideal bioscaffold for *in vivo* use must not induce a toxic or immune response or result in excessive inflammation. It should have an acceptably low (preferably no) level of disease risk, be slowly biodegradable, support the reconstruction of normal tissue and have similar mechanical and physical properties to the skin it replaces. It should also be repairable, readily available and capable of being prepared and stored with a long shelf life. Both natural materials (such as extracted FDA-approved bovine collagen or hyaluronic acid, or acellular natural human or porcine matrices) and synthetic FDA-approved biodegradable scaffolds (such as poly-L-lactide, used in dissolvable sutures and related polymers) are in use.

Cell delivery approaches

Improving ease of use for both clinicians and patients is essential if skin repair products are to achieve their full potential. A number of different strategies have been explored. Keratinocytes can be grown into integrated cell sheets (CEAs, now available as Epicel⁴³), made into suspensions and sprayed onto wound sites with⁴⁴ or without fibrin (for example, CellSpray⁴⁵) or delivered on various carrier dressings ranging from bovine collagen⁴⁶ to a chemically defined polymer carrier dressing (such as Myskin⁴⁷). Cells have also been cultured from hair follicles and delivered as small sheets (for example EpiDex⁴⁸). So far, there have been few studies comparing one method with another⁴⁹. Irrespective of how keratinocytes are delivered to the wound bed, healing success depends mostly on the condition of the wound bed — the condition of keratinocytes is less important. Consequently, the use of any one particular form of cell delivery will ultimately be determined by factors such as ease of use, cost, transportation and variables other than the biological health of the cells.

An example of this has been provided by work at Sheffield University in the UK, where cultured cells have been delivered to burns patients since 1992. A 10-year audit revealed that there was considerable waste during the use of CEA sheets, as this method does not allow much flexibility in the timing of cell delivery to the patient⁵⁰. Accordingly, the preference now is for subconfluent cells to be delivered on a chemically defined carrier dressing (for both chronic wounds⁵¹ and burns injuries⁵²), from which the cells transfer to the wound bed. The inert carrier dressing is coated with an extremely thin polymer containing acid functional groups using the technique of plasma polymerization⁵³. Repeated applications can promote wound healing in chronic wounds⁵¹. As with other such studies, this technique requires regular repeated applications of cells (either autologous or allogeneic) to achieve healing in wounds that may have been 'stuck' for a long period of time⁵⁴.

Clinical success but economic failure

The relatively small number of patients with severe burns in the developed world and the challenging economics of treating these patients explain why relatively few commercial companies have developed autologous keratinocyte-based products for use in this area. Most companies have focused on the larger chronic wounds market with more off-the-shelf approaches using either cell-free materials as temporary wound dressings (for example Integra²¹) or donor cells in some form of scaffold (such as Dermagraft²⁸). Tissue engineering of skin had such obvious clinical benefits that it led to unrealistic commercial and clinical expectations. Initially, the emphasis was solely on patient benefit, but it quickly turned to commercial viability. However, there was no clear framework for the regulation of tissue engineering, and the need for the parallel development of regulatory controls (a very slow process) in this new medical area has been a considerable problem for commercial operations. Considerable progress has been made, but after 25 years there is still no uniform international approach to this issue. Administrative lead times for these new activities have proved slower than scientific lead times. Tissue-engineered products do not fit healthcare models for developing medical devices or drugs.

Treatment of patients with extensive burns is neither cheap nor easy. The cost of production is always higher with autologous cells than with allogeneic cells. Although the latter have many uses, they cannot be used for the permanent repair of severe burns and financing these high costs of production was a challenge in the field's early years. Against this background, it is not surprising that some of the companies that have made major contributions to the area and developed clinically effective products have not achieved commercial success. For example, Dermagraft²⁸ achieved clinical success in treating non-healing ulcers but is not yet commercially viable. Others, such as Apligraf³⁰, have survived in the market place and been a financial success.

With hindsight, there are some obvious lessons. A new market is unlikely to reimburse high-cost products until these have been proved to be considerably more effective than standard treatments. The need for user-friendly products that do not require specialist expertise is clear. Finally, the speed with which regulatory mechanisms can be developed is a critical success factor, especially for smaller companies. The 'instant successes' reported by the popular press are, in reality, the result of stubborn persistence by relatively few groups.

Laboratory uses of tissue-engineered skin

In addition to the clinical uses of tissue-engineered skin there are many non-clinical research applications for reconstructed skin tissue, especially 3D models. Engineered skin is currently delivering value in aspects of skin biology research as diverse as reducing animal experimentation, investigation of cell-cell and cell-extracellular-matrix interactions, skin barrier penetration, wound healing, angiogenesis, regulation of pigmentation, skin contraction and investigation of skin diseases such as melanoma invasion, psoriasis and skin blistering disorders. The next decade will see increasing use of 3D models, in which physiological interactions between different skin cells that cannot occur in standard monolayer cultures are possible. Much current knowledge is based on 2D cultures, and this will need to be revalidated.

For *in vitro* use, the design criteria for 3D scaffolds are less exacting than for *in vivo* use. The scaffold should be fit for purpose, that is, it should allow skin cells to form an epithelium, to communicate with each other and to mimic as far as possible the responses of normal skin cells when exposed to a range of biological agents. The reconstructed skin should deliver the required biology, but, for example, for *in vitro* toxicity studies, scaffolds made of non-biodegradable material may be adequate^{55,56}. These scaffolds lack any functioning vasculature or immune system, so cannot be used as a complete substitute for *in vivo* animal or human studies, but they have a significant advantage in the ability to add or delete different cell types to assess their relevance to the aspect of skin biology under investigation. This cannot be readily done *in vivo*. These models are contributing to many areas of research, some of which are detailed below.

Alternatives to animal testing

One of the main impetuses for developing tissue-engineered skin has been increased pressure to find alternatives to animal testing for the many thousands of chemical additives used in human skin products. In Europe, humanitarian reasons for reducing animal experimentation have now translated to directives from the European Union⁵⁷. Although such models cannot completely remove the need for some animal experimentation (because of the lack of immune and circulatory responses) they can massively reduce the number of animal dermatotoxicity tests. Substantial investment in the development of skin models for testing pharmaceutical cosmetic and chemical compounds has been undertaken in recent years by companies such as L'Oreal^{58,59} and Skin Ethic⁶⁰. This has led to the development of epithelial cell models to study agents that might induce toxicity or irritancy. At present, two reconstructed human skin models, Epiderm and Episkin, are being evaluated in an ECVAM (European Centre for the Validation of Alternative Methods) Skin Irritation Validation Study⁶¹. However, although many commercial 3D models of skin contain keratinocytes, they need to be improved to include fibroblasts. Recent experience⁵⁶ shows that the presence of fibroblasts helps keratinocytes cope with potentially toxic agents. Fibroblast support

Box 1 | Examples of currently available skin-replacement materials

Epithelial cover

Involves the delivery of autologous keratinocytes as one of the following:

- An integrated sheet such as Epicel (Genzyme Tissue Repair). This is developed from the methodology originally pioneered in 1981 (ref. 12). A biopsy of the patient's cells is grown into an integrated sheet and enzymatically detached for delivery to the patient⁴³.
- Subconfluent cells on a carrier such as Myskin (CellTran)⁴⁷. Cells are delivered to the patient before they reach confluence on a chemically defined carrier dressing.
- Small sheets cultured from a patient's hair follicles such as Epidex (Modex Therapeutics)⁴⁸.
- A spray such as CellSpray (Clinical Cell Culture). Subconfluent cells are expanded in the laboratory and made into a suspension in which they are transported. They are then delivered to the patient as a spray⁴⁵.

Dermal replacement materials

- Donor skin³⁴: skin from screened skin donors can be used to provide either a temporary wound cover or a permanent source of allodermis.
- Integra²¹ (Integra LifeSciences): an alternative to donor skin that provides a vascularized dermis for a subsequent split-thickness skin graft.
- Alloderm (Lifecell): freeze-dried human donor dermis³⁵.
- Dermagraft (Advanced Biohealing): a synthetic material conditioned with donor fibroblasts²⁸.
- Transcyte (Advanced Biohealing): similar to Dermagraft but with a silicone membrane to act as a temporary epidermal barrier²⁹.
- Permacol (Tissue Science Laboratories): porcine skin that provides a temporary wound dressing³⁶.

Epidermal/dermal replacement materials

- Apligraf (Organogenesis): this combines allogeneic keratinocytes and fibroblasts with bovine collagen to provide a temporary skin-replacement material suitable for use in chronic wounds but not major burns³⁰.
- OrCel (Ortec International): combines allogeneic keratinocytes and fibroblasts with bovine collagen to provide a temporary skin-replacement material suitable for use in chronic wounds³¹.
- Cincinnati skin substitute, or Permaderm (Cambrex): comprises autologous keratinocytes and fibroblasts crafted into reconstructed skin with bovine collagen. Can provide a permanent skin substitute for burns patients³⁷.

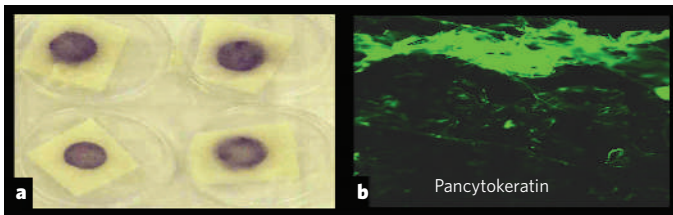


Figure 3 | Reconstruction of tissue-engineered skin using a synthetic scaffold. Shows reconstruction of 3D tissue-engineered skin for *in vitro* purposes only using an electrospun polystyrene scaffold. The polystyrene scaffold contained fibres of about 10 μm thickness and had a packing fraction of 0.1; it was produced by electrospinning. **a**, The presence of skin cells (keratinocytes, fibroblasts and endothelial cells) cultured in the scaffold for 10 days and revealed with an MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) viable stain. The area in which the cells were seeded was 1 cm in diameter. **b**, A cross-section through this scaffold — which has been seeded with keratinocytes, fibroblasts and endothelial cells — in which keratinocytes have been specifically identified by immunostaining with pancytokeratin. (Images reproduced, with permission, from ref. 55.)

of keratinocyte expansion and organization can be demonstrated even in a 3D electrospun polystyrene matrix that has none of the specialized collagen structure of the natural dermis⁵⁵ (Fig. 3).

Pigmentation

Excellent work on understanding the biology of both normal and vitiligo melanocytes in reconstructed skin has been published by the Taieb laboratory⁶². The ability to do 'mix and match' experiments led to findings that melanocytes in 3D culture spontaneously produce pigment reflecting their donor phototype origin. This ability was not influenced by the keratinocyte environment.

Following on from this, the L'Oreal group⁵⁸ have developed a model with melanocytes for investigating pigmentary responses to agents. Surprisingly, fibroblasts were found to have a substantial role in determining whether spontaneous pigmentation occurs in reconstructed skin⁶³. This is illustrated in Fig. 4a, b. Figure 4a shows a section through skin, indicating the location of the melanocytes in the lower layer of the epidermis. In Fig. 4b, however, the centre of the reconstructed skin discs was deliberately not seeded with melanocytes to provide contrast. In the presence of fibroblasts, repopulating reconstructed skin with melanocytes taken from pale-skinned donors resulted in amelanotic skin (Fig. 4b, left-hand model). Omission of the fibroblasts in the same experiments led to spontaneous pigmentation⁶³ (Fig. 4b, right-hand model). Recent work⁶⁴ in immune-tolerant mice grafted with reconstructed human skin has confirmed that human fibroblasts also function to reduce melanocyte pigmentation *in vivo*. Thus, melanocyte behaviour is very different in the presence and absence of fibroblasts — a result that can be easily seen with the appropriate combination of cells present in the 3D model but that is difficult to see in 2D. The role of fibroblasts in reducing pigmentation was not obvious before the development of 3D models.

It has been noted that hypopigmentation is often a problem when reconstructed skin is used for burns patients⁴¹. When collagen is used as the dermal scaffold on which to reconstruct skin, the melanocytes initially present in the tissue-engineered skin may be rapidly lost from these reconstructs after grafting⁴¹. This can be duplicated experimentally using reconstructed human skin that has been prepared with or without a basement membrane. In the absence of a basement membrane, melanocytes fail to make a secure attachment and are rapidly lost⁶³ (see page 843).

Melanoma invasion

3D reconstructed models have provided insight into the interaction between tumour cells and healthy skin cells. A number of research groups — such as that of Delvoye *et al.*⁶⁵ — are examining the cell interactions involved in early-stage melanoma invasion. Such models can be used to look at the extent of invasion by different melanoma cell lines,

to examine the deliberate upregulation or downregulation of adhesion molecules on these melanomas, and study matrix metalloproteinases and inflammation. Invasion (Fig. 4c) has been shown to be dependent on the interaction of the tumour cells with the adjacent skin cells^{66–68}.

Skin contraction

Unexpected results have been obtained in relation to the contraction of human skin grafts. The data from work using reconstructed skin from sterilized mature crosslinked dermis (Fig. 4d) show that keratinocytes rather than fibroblasts contract this dermis^{69–71}. Both keratinocytes⁷² and fibroblasts⁷³ can contract dilute collagen gels, but the finding that keratinocytes have a marked ability to contract human dermis sheds new light on the problem of skin-graft contracture. This model is being used to identify new pharmacological targets to block contraction. Recent evidence shows that inhibitors of the collagen crosslinking enzyme lysyl oxidase can block contraction to a large degree without having adverse effects on skin morphology⁷¹.

Other uses

Reconstructed skin has also been used to develop models of human skin diseases such as psoriasis⁷⁴ and the genetic disorder epidermolysis bullosa — a severe, often lethal blistering problem. Here it is being used to develop therapeutic gene expression to improve keratinocyte

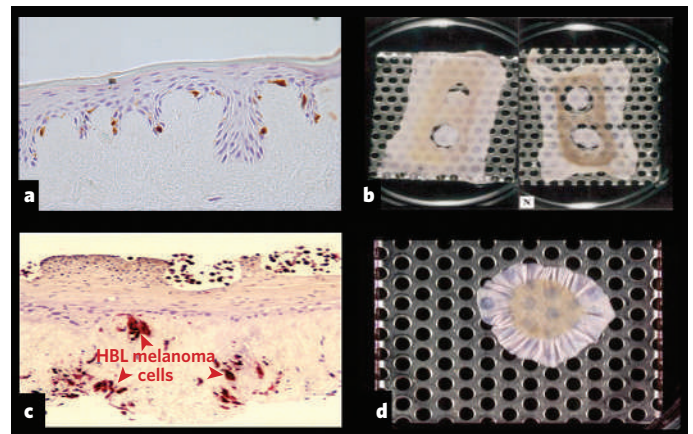


Figure 4 | Examples of laboratory uses of tissue-engineered skin. Reconstructed human skin based on de-epidermized acellular dermis has various uses — examples include the study of skin pigmentation, melanoma invasion and skin graft contraction. **a**, For skin pigmentation studies, melanocytes were added to reconstructed skin containing keratinocytes and fibroblasts, cultured for 10 days and then sectioned and stained with eosin and haematoxylin to show their general morphology. Melanocytes are identified by immunostaining for S100, and are distributed in the basal keratinocyte layer. In this study, keratinocytes have formed a well-attached epithelium with finger-like projections into the dermis that are known as rete ridges. **b**, Investigation of the contribution of fibroblasts to skin pigmentation. The melanocytes used were obtained from a pale-skinned donor and combined in a 5/95 melanocyte/keratinocyte ratio. Fibroblasts were included in the left-hand model and omitted from the right-hand model. The reconstructed skin models were cultured for 10 days, with two discs 8 mm in diameter removed after 5 days and replaced with material lacking any additional melanocytes for comparison. As can be seen, the omission of fibroblasts led to spontaneous pigmentation in this reconstructed skin model⁶³. **c**, To study melanoma invasion, the human melanoma cell line (HBL) was added to reconstructed human skin containing keratinocytes and fibroblasts, and a section through this reconstructed skin was stained with eosin and haematoxylin for general morphology. Melanoma cells were specifically identified by staining with the HMB45 antibody using a red endpoint (arrows). Penetration of melanoma cells through the epidermis into the dermis can be seen clearly⁶⁶. **d**, Demonstration of the use of this model to study skin graft contraction. De-epidermized acellular human dermis, to which only keratinocytes had been added, were cultured in an air–liquid interface for 10 days. The ability of keratinocytes to gather in the underlying dermis is evident^{69–71}.

adhesion. At present, there is little to offer these patients unless gene therapy and grafting of tissue-engineered skin can be achieved, and this is now looking promising⁷⁵.

Challenges for the future

Tissue engineering of skin is based on almost 25 years' worth of published research and rests on a strong bedrock of cell biology and wound-healing knowledge. There is a relatively mature understanding of what works well, what remains challenging and where to concentrate effort. At a prosaic level there remains a major need to reduce the risks for patients receiving cultured cells. For such patients, a reliable and regulatory-body-approved xenobiotic-free keratinocyte culture methodology is badly needed. Dermal substitute materials that present lower risks for burns patients than bovine collagen or human donor dermis are also needed. In terms of success after grafting, the main challenge is to get cultured cells to attach securely — often to very challenging wound beds. Means of accelerating basement membrane formation and vascularization are needed to help tissue-engineered skin survive after grafting. After this, graft contraction, loss of pigmentation and scarring at the margins may occur, as with conventional skin grafts. *In vitro* models of tissue-engineered skin are offering insight into how to prevent contraction and loss of pigmentation. With respect to scar prevention, a long-established programme of research^{76,77} on understanding the role of the transforming growth factor- β family in wound healing and scar formation is now being developed into a therapy⁷⁸.

The experience of the past 25 years has identified that in the production of tissue-engineered materials the focus must move more quickly from laboratory to clinical use. Products must be user-friendly. Production costs must be contained so that reimbursement gives a sufficient income. Commercial lessons have been hard won, but products have successfully negotiated a rapidly changing regulatory environment. Safe and clinically effective materials have been developed, and making these commercially viable is now feasible. Despite initial unrealistic commercial and clinical expectations, tissue-engineered skin has delivered considerable benefits to patients with burns and chronic wounds and has enormous potential that is only just beginning to be realized. ■

- Herndon, D. N. *et al.* A comparison of conservative versus early excision. Therapies in severely burned patients. *Ann. Surg.* **209**, 547–552 (1989).
- Burn Incidence and Treatment in the United States: 1999 Fact Sheet* (The Burn Foundation, Philadelphia, 1999).
- Rose, J. K. & Herndon, D. N. Advances in the treatment of burn patients. *Burns* **23** (suppl. 1), S19–S26 (1997).
- Falanga, V. Chronic wounds: pathophysiologic and experimental considerations. *J. Invest. Dermatol.* **100**, 721–725 (1993).
- Phillips, T., Stanton, B., Provan, A. & Lew, R. A study of the impact of leg ulcers on quality of life: financial, social, and psychologic implications. *J. Am. Acad. Dermatol.* **31**, 49–53 (1994).
- Bittencourt, F. V. *et al.* Large congenital melanocytic nevi and the risk of development of malignant melanoma and neurocutaneous melanocytosis. *Pediatrics* **106**, 736–741 (2000).
- Cooper, M. L., Spielvogel, R. L., Hansbrough, J. F., Boyce, S. T. & Frank, D. H. Reconstitution of the histologic characteristics of a giant congenital nevomelanocytic nevus employing the athymic mouse and a cultured skin substitute. *J. Invest. Dermatol.* **97**, 649–658 (1991).
- Gallico, G. G. *et al.* Cultured epithelial autografts for giant congenital nevi. *Plast. Reconstr. Surg.* **84**, 1–9 (1989).
- Rheinwald, J. G. & Green, H. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell* **6**, 331–343 (1975).
- Rheinwald, J. G. & Green, H. Epidermal growth factor and the multiplication of cultured human epidermal keratinocytes. *Nature* **265**, 421–424 (1977).
- Green, H., Kehinde, O. & Thomas, J. Growth of cultured human epidermal cells into multiple epithelia suitable for grafting. *Proc. Natl Acad. Sci. USA* **76**, 5665–5668 (1979).
- O'Connor, N. E., Mulliken, J. B., Banks-Schlegel, S., Kehinde, O. & Green, H. Grafting of burns with cultured epithelium prepared from autologous epidermal cells. *Lancet* **i**, 75–78 (1981).
- Gallico, G. G., O'Connor, N. E., Compton, C. C., Kehinde, O. & Green, H. Permanent coverage of large burn wounds with autologous cultured human epithelium. *New Engl. J. Med.* **311**, 448–451 (1984).
- Barrandon, Y. & Green, H. Three clonal types of keratinocyte with different capacities for multiplication. *Proc. Natl Acad. Sci. USA* **84**, 2302–2306 (1987).
- Fuchs, E. Epidermal differentiation: the bare essentials. *J. Cell Biol.* **111**, 2807–2814 (1990).
- Gambardella, L. & Barrandon, Y. The multifaceted adult epidermal stem cell. *Curr. Opin. Cell Biol.* **15**, 771–777 (2003).
- Cuono, C., Langdon, R. & McGuire, J. Use of cultured epidermal autografts and dermal allografts as skin replacement after burn injury. *Lancet* **i**, 1123–1124 (1986).
- Herndon, D. N. & Parks, D. H. Comparison of serial debridement and autografting and early massive excision with cadaver skin overlay in the treatment of large burns in children. *J. Trauma* **26**, 149–152 (1986).
- Burke, J. F., Yannas, I. V., Quinby, W. C., Bondoc, C. C. & Jung, W. K. Successful use of a physiologically acceptable artificial skin in the treatment of extensive burn injury. *Ann. Surg.* **194**, 413–428 (1981).
- Huang, Q., Dawson, R. A., Pegg, D. E., Kearney, J. N. & MacNeil, S. Use of peracetic acid to sterilise human donor skin for production of acellular matrices for clinical use. *Wound Repair Regen.* **12**, 276–287 (2004).
- Stern, R., McPherson, M. & Longaker, M. T. Histologic study of artificial skin used in the treatment of full thickness thermal injury. *J. Burn Care Rehabil.* **11**, 7–13 (1990).
- Boyce, S. T. *et al.* Comparative assessment of cultured skin substitutes and native skin autograft for treatment of full-thickness burns. *Ann. Surg.* **222**, 743–752 (1995).
- Boyce, S. T. *et al.* The 1999 Clinical Research Award. Cultured skin substitutes combined with Integra artificial skin to replace native skin autograft and allograft for closure of full-thickness burns. *J. Burn Care Rehabil.* **20**, 453–461 (1999).
- Boyce, S. T. *et al.* Cultured skin substitutes reduce donor skin harvesting for closure of excised, full-thickness burns. *Ann. Surg.* **235**, 269–279 (2002).
- Supp, D. M. & Boyce, S. T. Engineered skin substitutes: practices and potentials. *Clin. Dermatol.* **23**, 403–412 (2005).
- Iuchi, S., Dabelsteen, S., Easley, K., Rheinwald, J. G. & Green, H. Immortalized keratinocyte lines derived from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **103**, 1792–1797 (2006).
- McPherson, T. B. *et al.* Gal(1,3)Gal epitope in porcine small intestinal submucosa. *Tissue Eng.* **6**, 233–239 (2000).
- Marston, W. A., Hanft, J., Norwood, P. & Pollak, R. The efficacy and safety of Dermagraft in improving the healing of chronic diabetic foot ulcers: results of a prospective randomized trial. *Diabetes Care* **26**, 1701–1705 (2003).
- Kumar, R. J., Kimble, R. M., Boots, R. & Pegg, S. P. Treatment of partial-thickness burns: a prospective randomized trial using Transcyte™. *ANZ J. Surg.* **74**, 622–626 (2004).
- Bello, Y. M. & Falabella, A. F. The role of Graftskin (Apligraf®) in difficult-to-heal venous leg ulcers. *J. Wound Care* **11**, 182–183 (2003).
- Lipkin, S., Chaikof, E., Isseroff, Z. & Silverstein, P. Effectiveness of bilayered cellular matrix in healing of neuropathic diabetic foot ulcers: results of a multicenter pilot trial. *Wounds* **15**, 230–236 (2003).
- Carver, N., Navsaria, H., Green, C. J. & Leigh, I. M. Acute rejection of cultured keratinocyte allografts in nonimmunosuppressed pigs. *Transplantation* **52**, 918–921 (1991).
- Brain, A. *et al.* Survival of cultured allogeneic keratinocytes transplanted to deep dermal bed assessed with probe specific for Y chromosome. *Br. Med. J.* **298**, 917–919 (1989).
- Hermans, M. H. Clinical experience with glycerol-preserved donor skin treatment in partial thickness burns. *Burns Incl. Therm. Inj.* **15**, 57–59 (1989).
- Wainwright, D. *et al.* Clinical evaluation of an acellular allograft dermal matrix in full-thickness burns. *J. Burn Care Rehabil.* **17**, 124–136 (1996).
- Jarman-Smith, M. L. *et al.* Porcine collagen crosslinking, degradation and its capability for fibroblast adhesion and proliferation. *J. Mater. Sci. Mater. Med.* **15**, 925–932 (2004).
- Boyce, S. T. *et al.* Cultured skin substitutes reduce requirements for harvesting of skin autograft for closure of excised, full-thickness burns. *J. Trauma* **60**, 821–829 (2006).
- Ralston, D. R. *et al.* The requirement for basement membrane antigens in the production of human epidermal/dermal composites *in vitro*. *Br. J. Dermatol.* **140**, 605–615 (1999).
- Sahota, P. S. *et al.* Development of a reconstructed human skin model for angiogenesis. *Wound Repair Regen.* **11**, 275–284 (2003).
- Medalie, D. A. *et al.* Differences in dermal analogs influence subsequent pigmentation, epidermal differentiation, basement membrane and rete ridge formation of transplanted composite skin grafts. *Transplantation* **64**, 454–465 (1997).
- Swope, V. B., Supp, A. P., Schwemmer, S., Babcock, G. & Boyce, S. Increased expression of integrins and decreased apoptosis correlate with increased melanocyte retention in cultured skin substitutes. *Pigment Cell Res.* **19**, 424–433 (2006).
- Ratner, B. D. & Bryant, S. J. Biomaterials: where we have been and where we are going. *Annu. Rev. Biomed. Eng.* **6**, 41–75 (2004).
- Wright, K. A. *et al.* Alternative delivery of keratinocytes using a polyurethane membrane and the implications for its use in the treatment of full-thickness burn injury. *Burns* **24**, 7–17 (1998).
- Horch, R. E., Bannasch, H. & Stark, G. B. Transplantation of cultured autologous keratinocytes in fibrin sealant biomatrix to resurface chronic wounds. *Transplant Proc.* **33**, 642–644 (2001).
- Navarro, F. A. *et al.* Sprayed keratinocyte suspensions accelerate epidermal coverage in a porcine microwound model. *J. Burn Care Rehabil.* **21**, 513–518 (2000).
- Horch, R. E., Debus, M., Wagner, G. & Stark, G. B. Cultured human keratinocytes on type I collagen membranes to reconstitute the epidermis. *Tissue Eng.* **6**, 53–67 (2000).
- Haddow, D. B. *et al.* Plasma-polymerized surfaces for culture of human keratinocytes and transfer of cells to an *in vitro* wound-bed model. *J. Biomed. Mater. Res. A* **64**, 80–87 (2003).
- Tausche, A. K. *et al.* An autologous epidermal equivalent tissue-engineered from follicular outer root sheath keratinocytes is as effective as split-thickness skin autograft in recalcitrant vascular leg ulcers. *Wound Repair Regen.* **11**, 248–252 (2003).
- Currie, L. J., Martin, R., Sharpe, J. R. & James, S. E. A comparison of keratinocyte cell sprays with and without fibrin glue. *Burns* **29**, 677–685 (2003).
- Hernon, C. A. *et al.* Clinical experience using cultured epithelial autografts lead to an alternative methodology for transferring skin cells from the laboratory to the patient. *Regen. Med.* **1**, 809–821 (2006).
- Moustafa, M. *et al.* A new autologous keratinocyte dressing treatment for non-healing diabetic neuropathic foot ulcers. *Diabet. Med.* **21**, 786–789 (2004).
- Zhu, N. *et al.* Treatment of burns and chronic wounds using a new cell transfer dressing for delivery of autologous keratinocytes. *Eur. J. Plast. Surg.* **28**, 319–330 (2005).
- France, R. M., Short, R. D., Dawson, R. A. & MacNeil, S. Attachment of human keratinocytes to plasma co-polymers of acrylic acid/octa-1,7-diene and allyl amine/octa-1,7-diene. *J. Mater. Chem.* **8**, 37–42 (1998).
- De Luca, M. *et al.* Multicentre experience in the treatment of burns with autologous and allogeneic cultured epithelium, fresh or preserved in a frozen state. *Burns* **15**, 303–309 (1989).
- Sun, T. *et al.* Self-organisation of skin cells in three-dimensional electrospun polystyrene scaffolds. *Tissue Eng.* **11**, 1023–1033 (2005).

56. Sun, T., Jackson, S., Haycock, J. W. & MacNeil, S. Culture of skin cells in 3D rather than 2D improves their ability to survive exposure to cytotoxic agents. *J. Biotechnol.* **122**, 372–381 (2006).
57. Sauer, U. G., Spielmann, H. & Rusche, B. Fourth EU report on the statistics on the number of animals used for scientific purposes in 2002 — trends, problems, conclusions. *ALTEX* **22**, 59–67 (2002).
58. Facy, V., Flouret, V., Regnier, M. & Schmidt, R. Reactivity of Langerhans cells in human reconstructed epidermis to known allergens and UV radiation. *Toxicol. In Vitro* **19**, 787–795 (2005).
59. Welss, T., Basketter, D. A. & Schroder, K. R. *In vitro* skin irritation: facts and future. State of the art review of mechanisms and models. *Toxicol. In Vitro* **18**, 231–243 (2004).
60. Netzlaff, F., Lehr, C. M., Wertz, P. W. & Schaefer, U. F. The human epidermis models EpiSkin, SkinEthic and EpiDerm: an evaluation of morphology and their suitability for testing phototoxicity, irritancy, corrosivity, and substance transport. *Eur. J. Pharm. Biopharm.* **60**, 167–178 (2005).
61. Kandarova, H. *et al.* Assessment of the skin irritation potential of chemicals by using the SkinEthic reconstructed human epidermal model and the common skin irritation protocol evaluated in the ECVAM skin irritation validation study. *Altern. Lab. Anim.* **34**, 393–406 (2006).
62. Bessou, S. *et al.* *Ex vivo* study of skin phototypes. *J. Invest. Dermatol.* **107**, 684–688 (1996).
63. Hedley, S. J. *et al.* Fibroblasts play a regulatory role in the control of pigmentation in reconstructed human skin from skin types I and II. *Pigment Cell Res.* **15**, 49–56 (2002).
64. Cario-Andre, M., Pain, C., Gauthier, Y., Casoli, V. & Taieb, A. *In vivo* and *in vitro* evidence of dermal fibroblasts influence on human epidermal pigmentation. *Pigment Cell Res.* **19**, 434–442 (2006).
65. Meier, F., Nesbit, M. & Hsu, M.-Y. Human melanoma progression in skin reconstructs: biological significance of bFGF. *Am. J. Pathol.* **156**, 193–200 (2000).
66. Eves, P. *et al.* Characterization of an *in vitro* model of human melanoma invasion based on reconstructed human skin. *Br. J. Dermatol.* **142**, 210–222 (2003).
67. Eves, P. *et al.* Melanoma invasion in reconstructed human skin is influenced by skin cells — investigation of the role of proteolytic enzymes. *Clin. Exp. Metastasis* **20**, 685–700 (2003).
68. Eves, P. *et al.* Anti-inflammatory and anti-invasive effects of alpha-melanocyte-stimulating hormone in human melanoma cells. *Br. J. Cancer* **89**, 2004–2015 (2003).
69. Ralston, D. R. *et al.* Keratinocytes contract normal human dermal extracellular matrix and reduce soluble fibronectin production by fibroblasts in a skin composite model. *Br. J. Plast. Surg.* **50**, 408–415 (1997).
70. Chakrabarty, K. H. *et al.* Keratinocyte-driven contraction of reconstructed human skin. *Wound Repair Regen.* **9**, 95–106 (2001).
71. Harrison, C. A. *et al.* Use of an *in vitro* model of tissue-engineered human skin to investigate the mechanism of skin graft contraction. *Tissue Eng.* 1 Oct 2006 [Epub ahead of print].
72. Souren, J. M., Ponc, M. & Van Wijk, R. Contraction of collagen by human fibroblasts and keratinocytes. *In Vitro Cell Dev. Biol.* **25**, 1039–1045 (1989).
73. Grinnell, F. Fibroblasts, myofibroblasts and wound contraction. *J. Cell Biol.* **124**, 401–404 (1994).
74. Barker, C. L. *et al.* The development and characterization of an *in vitro* model of psoriasis. *J. Invest. Dermatol.* **123**, 892–901 (2004).
75. Ferrari, S., Pellegrini, G., Matsui, T., Mavilio, F. & De Luca, M. Gene therapy in combination with tissue engineering to treat epidermolysis bullosa. *Expert Opin. Biol. Ther.* **6**, 367–378 (2006).
76. Shah, M. *et al.* Role of elevated plasma transforming growth factor β 1 levels in wound healing. *Am. J. Pathol.* **154**, 1115–1124 (1989).
77. O'Kane, S. & Ferguson, M. W. Transforming growth factor β s and wound healing. *Int. J. Biochem. Cell Biol.* **29**, 63–78 (1997).
78. Ferguson, M. W. & O'Kane, S. Scar-free healing: from embryonic mechanisms to adult therapeutic intervention. *Phil. Trans. R. Soc. Lond. B* **359**, 839–850 (2004).
79. Shier, D., Butler, J. & Lewis, R. in *Hole's Human Anatomy and Physiology* 8th Edn 160–183 (McGraw Hill, 1999).
80. Herndon, D. N. & Parks, D. H. Comparison of serial debridement and autografting and early massive excision with cadaver skin overlay in the treatment of large burns in children. *J. Trauma* **26**, 149–152 (1986).
81. Delvoye, P. *et al.* Fibroblasts induce the assembly of the macromolecules of the basement membrane. *J. Invest. Dermatol.* **90**, 276–282 (1990).
82. Konig, A. & Brucker-Tuderman, L. Epithelial-mesenchymal interactions enhance expression of collagen VII *in vitro*. *J. Invest. Dermatol.* **96**, 803–808 (1991).
83. Demarchez, M., Hartmann, D. J., Regnier, M. & Asselineau, D. The role of fibroblasts in dermal vascularization and remodeling of reconstructed human skin after transplantation onto the nude mouse. *Transplantation* **54**, 317–326 (1992).

Acknowledgements The author regrets that space restrictions preclude full reference to all those who have contributed to this field. Sheila MacNeil's research is supported by grants from EPSRC, BBSRC and the Wellcome Trust.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to the author (s.macneil@shef.ac.uk).

BONE METASTASIS

Can osteoclasts be excluded?

Arising from: D. H. Jones *et al.* *Nature* **440**, 692–696 (2006)

The RANK/RANKL signalling mechanism is the final common pathway of osteoclast formation and activity¹. Inhibitors of RANK ligand (RANKL) that bind to RANK (for 'receptor activator of NF- κ B'), such as osteoprotegerin (OPG), neutralizing antibodies against RANKL and soluble RANK antagonists, are well described inhibitors of bone metastasis in preclinical and clinical models, presumably because of their effects on osteoclasts². Jones *et al.*³ show that OPG inhibits bone metastasis after intracardiac injection of B16F10 murine melanoma cells, but claim that bone metastases are entirely independent of osteoclast formation and bone resorption: rather, they are caused by an effect on cell migration through RANK. However, we question whether these surprising conclusions are rigorously supported by their data³.

The significance of RANK production by cancer cells has long been of interest, particularly as this TNF (for tumour-necrosis factor)-receptor family member is expressed in several breast-cancer cell lines and in a series of clinical breast cancers⁴. However, its role, if any, in the bone metastatic process is difficult to unravel in the absence of osteoclastic bone resorption. With this preclinical model of bone metastasis⁵, tumour burden at bone metastatic sites is decreased whenever osteoclasts are inhibited — whether by bisphosphonates, inhibition of

RANKL, neutralizing antibodies against parathyroid-hormone-related protein (PTHrP) or inhibitors of PTHrP transcription^{2,6,7}. Evidence for a central role for osteoclasts in bone metastasis is therefore compelling.

Jones *et al.*³ assume that osteoclasts have no role in their model of metastasis. To prove that, quantitative histology would be required, with the number of osteoclasts counted from the early stages of tumour-cell growth in bone. Once bone tumours have become large and generally invasive (which includes cell migration), it becomes difficult to see and quantify osteoclasts, even though classical large resorption-site Howship's lacunae are present. These lacunae are characteristic of osteoclasts and cannot be generated by any other cell, as shown in a scanning electron microscope study of human cancers metastasized to bone⁸. Jones *et al.*³ indicate (their data are not shown) that treatment with the bisphosphonate zoledronic acid is ineffective in preventing bone-tumour growth. However, they do not provide the doses used, the timing of treatment, evidence that the zoledronic acid can inhibit osteoclasts, details of the behaviour of appropriate tumour controls (such as breast-cancer cells), or indication of whether treatment with OPG in the same experiment is effective.

Investigations into the role of the bone microenvironment in helping cancers to

establish and grow in bone indicate that the attachment, maturation and activation of osteoclast precursors are likely to be important for tumour expansion, in which case the RANKL/RANK signalling system would be a suitable target for prevention and treatment. If the additional role of RANK-mediated cell motility proposed by Jones *et al.*³ does indeed replace that of osteoclasts in the case of B16F10 melanoma cells, the provision of more convincing evidence would dispel any confusion.

T. John Martin*, **Gregory R. Mundy†**

*St Vincent's Institute of Medical Research, Fitzroy, Victoria 3065, Australia
e-mail: jmartin@svi.edu.au

†Vanderbilt Centre for Bone Biology, 1235 Medical Research Building IV, Nashville, Tennessee 37232-0575, USA

1. Teitelbaum, S. L. & Ross, F. P. *Nature Rev. Genet.* **4**, 638–649 (2003).
2. Mundy, G. R. *Nature Rev. Cancer* **2**, 584–593 (2002).
3. Jones, D. H. *et al.* *Nature* **440**, 692–696 (2006).
4. Thomas, R. J. *et al.* *Endocrinology* **140**, 4451–4458 (1999).
5. Arguello, F., Baggs, R. B. & Frantz, C. N. *Cancer Res.* **48**, 6876–6881 (1988).
6. Guise, T. A. *et al.* *J. Clin. Invest.* **98**, 1544–1549 (1996).
7. Gallwitz, W. E., Guise, T. A. & Mundy, G. R. *J. Clin. Invest.* **110**, 1559–1572 (2002).
8. Boyde, A., Maconacchie, E., Reid, S. A., Delling, G. & Mundy, G. R. *Scanning Electron. Microsc.* **IV**, 1537–1554 (1986).

Competing financial interests: declared none.
doi:10.1038/nature05657

BONE METASTASIS

Jones *et al.* replyReplying to: T. J. Martin & G. R. Mundy *Nature* **445**, doi: 10.1038/nature05657 (2007)

Martin and Mundy¹ emphasize the crucial role played by osteoclasts in the metastasis of many tumour types to bone. Our intention was not to question the importance of osteoclasts in bone metastases, but rather to show that RANK ligand (RANKL) exerts additional action on cancer cells beyond its effects on osteoclasts². We reported that epithelial tumour cells can express RANK, a finding that is supported by earlier work³; that RANKL has chemotactic activity⁴; and that *in vivo* inhibition of RANKL by osteoprotegerin reduces tumour burden in bone when assessed in a previously published experimental system⁵.

Mundy and Martin¹ suggest that we cannot exclude the involvement of osteoclasts. For our *in vivo* studies, we chose a melanoma cell line that had previously been shown to metastasize to bone but not to trigger osteoclast activation⁵. We confirmed that spreading of these melanoma

cells to bone does not result in obvious activation of osteoclasts by using several complementary approaches²: histomorphometry of bone structure; measurements of serum calcium, phosphorus, alkaline phosphatase and tartrate-resistant acid phosphatase; and radiographic analyses of bone. As these methods are the same as those used by Mundy and colleagues in their studies on bone metastases^{6,7}, our methodology would seem to be sufficient. Contrary to Mundy and Martin's assertion, details of the doses of zoledronic acid we used and the timing of treatment are provided in our Supplementary Information (page 6)².

Differences of opinion aside, Mundy and Martin pose an important question: what is the role of osteoclasts in bone metastases? All of the proposed theories are based on correlative localization studies and experiments with bisphosphonates that affect not only osteoclasts

but also angiogenesis, apoptosis, proliferation and migration of other cell types^{7–9}. This question can ultimately be answered only by the use of genetic models. Such experiments were not possible, however, as all mutant mice with disrupted osteoclasts have severely altered bone structures, or have other pathways that have been affected. To answer this question genetically, we have generated RANK^{flxed} mice as a tool to knock out RANK in adult osteoclasts selectively, which should enable the role of osteoclasts in cancer metastases to be re-evaluated.

Our findings² describe an additional mechanism through which RANKL could promote bone metastases, namely the stimulation of cancer-cell migration. RANKL is the master gene for bone turnover through osteoclasts¹⁰ and has a key function in epithelial-cell proliferation in the mammary gland¹¹. Based

on our own and other evidence^{2–4}, including previous work by Martin and Mundy, RANKL qualifies as a bona fide soil factor that helps to explain the preferential metastasis of tumour cells into bones. The role of RANKL and RANK in cancer might extend beyond their function in osteoclasts and cell migration, to growth of the primary cancer. Because of its role in osteoclasts and in tumour cells, as also independently confirmed^{1–4,10,11}, inhibition of RANKL–RANK interaction offers a target for interfering with tumour metastasis and progression in bone.

D. Holstead Jones*†‡§, **Tomoki Nakashima**†, **Otto H. Sanchez**¶, **Ivona Kozieradzki**†‡§, **Svetlana V. Komarova**||, **Ildiko Sarosi**#, **Sean Morony**#, **Evelyn Rubin**‡§, **Renu Sarao**†, **Carlo V. Hojilla**¶, **Vukoslav Komnenovic**†, **Young-Yun Kong****†, **Martin Schreiber**††, **S. Jeffrey Dixon**‡‡, **Stephen M. Sims**‡‡, **Rama Khokha**‡¶, **Teiji Wada**†, **Josef M. Penninger**†‡§

*Faculty of Health Sciences, University of Ontario Institute of Technology, Oshawa, Ontario L1H 7K4, Canada

†Institute of Molecular Biotechnology of the Austrian Academy of Sciences, 1030 Vienna, Austria

e-mail: josef.penninger@imb.oeaw.ac.at

‡Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2C1, Canada

§Department of Immunology, University of Toronto, Toronto, Ontario M5G 2C1, Canada

¶Ontario Cancer Institute, University Health Network, Toronto, Ontario M5G 2C1, Canada

||Faculty of Dentistry, McGill University, Montreal, Quebec H3A 1A4, Canada

#Amgen Inc., Thousand Oaks, California 91320-1799, USA

**Division of Molecular and Life Sciences, Pohang University of Science and Technology, Pohang, 790-784 Kyungbuk, South Korea

††Medical University of Vienna, Department of Obstetrics and Gynecology, Waehringer Guertel

18-20, 1090 Vienna, Austria

‡‡CIHR Group in Skeletal Development and Remodeling, Department of Physiology and Pharmacology and Division of Oral Biology, Schulich School of Medicine & Dentistry, The University of Western Ontario, London, Ontario N6A 5C1, Canada

1. Mundy, T. J. & Mundy, G. R. *Nature* **445**, doi:10.1038/nature05657 (2007).
2. Jones, D. H. *et al. Nature* **440**, 692–696 (2006).
3. Thomas, R. J. *et al. Endocrinology* **140**, 4451–4458 (1999).
4. Mosheimer, B. A., Kaneider, N. C., Feistritz, C., Sturn, D. H. & Wiedermann, C. J. *Arthr. Rheum.* **50**, 2309–2316 (2004).
5. Sanchez-Sweetman, O. H., Lee, J., Orr, F. W. & Singh, G. *Eur. J. Cancer* **33**, 918–925 (1997).
6. Sterling, J. A. *et al. Cancer Res.* **66**, 7548–7553 (2006).
7. Hiraga, H., Williams, P. J., Mundy, G. R. & Yoneda, T. *Cancer Res.* **61**, 4418–4424 (2001).
8. Clezardin, P. *Cancer Treat. Rev.* **31** (suppl), 1–8 (2005).
9. Wood, J. *et al. J. Pharmacol. Exp. Ther.* **302**, 1055–1061 (2002).
10. Kong, Y.-Y. *et al. Nature* **397**, 315–323 (1999).
11. Fata, J. E. *et al. Cell* **103**, 41–50 (2000).

doi:10.1038/nature05658

BONE METASTASIS

Can osteoclasts be excluded?

Arising from: D. H. Jones *et al.* *Nature* **440**, 692–696 (2006)

The RANK/RANKL signalling mechanism is the final common pathway of osteoclast formation and activity¹. Inhibitors of RANK ligand (RANKL) that bind to RANK (for 'receptor activator of NF- κ B'), such as osteoprotegerin (OPG), neutralizing antibodies against RANKL and soluble RANK antagonists, are well described inhibitors of bone metastasis in preclinical and clinical models, presumably because of their effects on osteoclasts². Jones *et al.*³ show that OPG inhibits bone metastasis after intracardiac injection of B16F10 murine melanoma cells, but claim that bone metastases are entirely independent of osteoclast formation and bone resorption: rather, they are caused by an effect on cell migration through RANK. However, we question whether these surprising conclusions are rigorously supported by their data³.

The significance of RANK production by cancer cells has long been of interest, particularly as this TNF (for tumour-necrosis factor)-receptor family member is expressed in several breast-cancer cell lines and in a series of clinical breast cancers⁴. However, its role, if any, in the bone metastatic process is difficult to unravel in the absence of osteoclastic bone resorption. With this preclinical model of bone metastasis⁵, tumour burden at bone metastatic sites is decreased whenever osteoclasts are inhibited — whether by bisphosphonates, inhibition of

RANKL, neutralizing antibodies against parathyroid-hormone-related protein (PTHrP) or inhibitors of PTHrP transcription^{2,6,7}. Evidence for a central role for osteoclasts in bone metastasis is therefore compelling.

Jones *et al.*³ assume that osteoclasts have no role in their model of metastasis. To prove that, quantitative histology would be required, with the number of osteoclasts counted from the early stages of tumour-cell growth in bone. Once bone tumours have become large and generally invasive (which includes cell migration), it becomes difficult to see and quantify osteoclasts, even though classical large resorption-site Howship's lacunae are present. These lacunae are characteristic of osteoclasts and cannot be generated by any other cell, as shown in a scanning electron microscope study of human cancers metastasized to bone⁸. Jones *et al.*³ indicate (their data are not shown) that treatment with the bisphosphonate zoledronic acid is ineffective in preventing bone-tumour growth. However, they do not provide the doses used, the timing of treatment, evidence that the zoledronic acid can inhibit osteoclasts, details of the behaviour of appropriate tumour controls (such as breast-cancer cells), or indication of whether treatment with OPG in the same experiment is effective.

Investigations into the role of the bone microenvironment in helping cancers to

establish and grow in bone indicate that the attachment, maturation and activation of osteoclast precursors are likely to be important for tumour expansion, in which case the RANKL/RANK signalling system would be a suitable target for prevention and treatment. If the additional role of RANK-mediated cell motility proposed by Jones *et al.*³ does indeed replace that of osteoclasts in the case of B16F10 melanoma cells, the provision of more convincing evidence would dispel any confusion.

T. John Martin*, **Gregory R. Mundy†**

*St Vincent's Institute of Medical Research, Fitzroy, Victoria 3065, Australia
e-mail: jmartin@svi.edu.au

†Vanderbilt Centre for Bone Biology, 1235 Medical Research Building IV, Nashville, Tennessee 37232-0575, USA

1. Teitelbaum, S. L. & Ross, F. P. *Nature Rev. Genet.* **4**, 638–649 (2003).
2. Mundy, G. R. *Nature Rev. Cancer* **2**, 584–593 (2002).
3. Jones, D. H. *et al.* *Nature* **440**, 692–696 (2006).
4. Thomas, R. J. *et al.* *Endocrinology* **140**, 4451–4458 (1999).
5. Arguella, F., Baggs, R. B. & Frantz, C. N. *Cancer Res.* **48**, 6876–6881 (1988).
6. Guise, T. A. *et al.* *J. Clin. Invest.* **98**, 1544–1549 (1996).
7. Gallwitz, W. E., Guise, T. A. & Mundy, G. R. *J. Clin. Invest.* **110**, 1559–1572 (2002).
8. Boyde, A., Maconacchie, E., Reid, S. A., Delling, G. & Mundy, G. R. *Scanning Electron. Microsc.* **IV**, 1537–1554 (1986).

Competing financial interests: declared none.
doi:10.1038/nature05657

BONE METASTASIS

Jones *et al.* replyReplying to: T. J. Martin & G. R. Mundy *Nature* **445**, doi: 10.1038/nature05657 (2007)

Martin and Mundy¹ emphasize the crucial role played by osteoclasts in the metastasis of many tumour types to bone. Our intention was not to question the importance of osteoclasts in bone metastases, but rather to show that RANK ligand (RANKL) exerts additional action on cancer cells beyond its effects on osteoclasts². We reported that epithelial tumour cells can express RANK, a finding that is supported by earlier work³; that RANKL has chemotactic activity⁴; and that *in vivo* inhibition of RANKL by osteoprotegerin reduces tumour burden in bone when assessed in a previously published experimental system⁵.

Mundy and Martin¹ suggest that we cannot exclude the involvement of osteoclasts. For our *in vivo* studies, we chose a melanoma cell line that had previously been shown to metastasize to bone but not to trigger osteoclast activation⁵. We confirmed that spreading of these melanoma

cells to bone does not result in obvious activation of osteoclasts by using several complementary approaches²: histomorphometry of bone structure; measurements of serum calcium, phosphorus, alkaline phosphatase and tartrate-resistant acid phosphatase; and radiographic analyses of bone. As these methods are the same as those used by Mundy and colleagues in their studies on bone metastases^{6,7}, our methodology would seem to be sufficient. Contrary to Mundy and Martin's assertion, details of the doses of zoledronic acid we used and the timing of treatment are provided in our Supplementary Information (page 6)².

Differences of opinion aside, Mundy and Martin pose an important question: what is the role of osteoclasts in bone metastases? All of the proposed theories are based on correlative localization studies and experiments with bisphosphonates that affect not only osteoclasts

but also angiogenesis, apoptosis, proliferation and migration of other cell types^{7–9}. This question can ultimately be answered only by the use of genetic models. Such experiments were not possible, however, as all mutant mice with disrupted osteoclasts have severely altered bone structures, or have other pathways that have been affected. To answer this question genetically, we have generated RANK^{flxed} mice as a tool to knock out RANK in adult osteoclasts selectively, which should enable the role of osteoclasts in cancer metastases to be re-evaluated.

Our findings² describe an additional mechanism through which RANKL could promote bone metastases, namely the stimulation of cancer-cell migration. RANKL is the master gene for bone turnover through osteoclasts¹⁰ and has a key function in epithelial-cell proliferation in the mammary gland¹¹. Based

on our own and other evidence^{2–4}, including previous work by Martin and Mundy, RANKL qualifies as a bona fide soil factor that helps to explain the preferential metastasis of tumour cells into bones. The role of RANKL and RANK in cancer might extend beyond their function in osteoclasts and cell migration, to growth of the primary cancer. Because of its role in osteoclasts and in tumour cells, as also independently confirmed^{1–4,10,11}, inhibition of RANKL–RANK interaction offers a target for interfering with tumour metastasis and progression in bone.

D. Holstead Jones*†‡§, **Tomoki Nakashima**†, **Otto H. Sanchez**¶, **Ivona Kozieradzki**†‡§, **Svetlana V. Komarova**||, **Ildiko Sarosi**#, **Sean Morony**#, **Evelyn Rubin**‡§, **Renu Sarao**†, **Carlo V. Hojilla**¶, **Vukoslav Komnenovic**†, **Young-Yun Kong****†, **Martin Schreiber**††, **S. Jeffrey Dixon**‡‡, **Stephen M. Sims**‡‡, **Rama Khokha**‡¶, **Teiji Wada**†, **Josef M. Penninger**†‡§

*Faculty of Health Sciences, University of Ontario Institute of Technology, Oshawa, Ontario L1H 7K4, Canada

†Institute of Molecular Biotechnology of the Austrian Academy of Sciences, 1030 Vienna, Austria

e-mail: josef.penninger@imb.oeaw.ac.at

‡Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2C1, Canada

§Department of Immunology, University of Toronto, Toronto, Ontario M5G 2C1, Canada

¶Ontario Cancer Institute, University Health Network, Toronto, Ontario M5G 2C1, Canada

||Faculty of Dentistry, McGill University, Montreal, Quebec H3A 1A4, Canada

#Amgen Inc., Thousand Oaks, California 91320-1799, USA

**Division of Molecular and Life Sciences, Pohang University of Science and Technology, Pohang, 790-784 Kyungbuk, South Korea

††Medical University of Vienna, Department of Obstetrics and Gynecology, Waehringer Guertel

18-20, 1090 Vienna, Austria

‡‡CIHR Group in Skeletal Development and Remodeling, Department of Physiology and Pharmacology and Division of Oral Biology, Schulich School of Medicine & Dentistry, The University of Western Ontario, London, Ontario N6A 5C1, Canada

1. Mundy, T. J. & Mundy, G. R. *Nature* **445**, doi:10.1038/nature05657 (2007).
2. Jones, D. H. *et al. Nature* **440**, 692–696 (2006).
3. Thomas, R. J. *et al. Endocrinology* **140**, 4451–4458 (1999).
4. Mosheimer, B. A., Kaneider, N. C., Feistritz, C., Sturn, D. H. & Wiedermann, C. J. *Arthr. Rheum.* **50**, 2309–2316 (2004).
5. Sanchez-Sweetman, O. H., Lee, J., Orr, F. W. & Singh, G. *Eur. J. Cancer* **33**, 918–925 (1997).
6. Sterling, J. A. *et al. Cancer Res.* **66**, 7548–7553 (2006).
7. Hiraga, H., Williams, P. J., Mundy, G. R. & Yoneda, T. *Cancer Res.* **61**, 4418–4424 (2001).
8. Clezardin, P. *Cancer Treat. Rev.* **31** (suppl), 1–8 (2005).
9. Wood, J. *et al. J. Pharmacol. Exp. Ther.* **302**, 1055–1061 (2002).
10. Kong, Y.-Y. *et al. Nature* **397**, 315–323 (1999).
11. Fata, J. E. *et al. Cell* **103**, 41–50 (2000).

doi:10.1038/nature05658

A genome-wide association study identifies novel risk loci for type 2 diabetes

Robert Sladek^{1,2,4}, Ghislain Rocheleau^{1*}, Johan Rung^{4*}, Christian Dina^{5*}, Lishuang Shen¹, David Serre¹, Philippe Boutin⁵, Daniel Vincent⁴, Alexandre Belisle⁴, Samy Hadjadj⁶, Beverley Balkau⁷, Barbara Heude⁷, Guillaume Charpentier⁸, Thomas J. Hudson^{4,9}, Alexandre Montpetit⁴, Alexey V. Pshezhetsky¹⁰, Marc Prentki^{10,11}, Barry I. Posner^{2,12}, David J. Balding¹³, David Meyre⁵, Constantin Polychronakos^{1,3} & Philippe Froguel^{5,14}

Type 2 diabetes mellitus results from the interaction of environmental factors with a combination of genetic variants, most of which were hitherto unknown. A systematic search for these variants was recently made possible by the development of high-density arrays that permit the genotyping of hundreds of thousands of polymorphisms. We tested 392,935 single-nucleotide polymorphisms in a French case-control cohort. Markers with the most significant difference in genotype frequencies between cases of type 2 diabetes and controls were fast-tracked for testing in a second cohort. This identified four loci containing variants that confer type 2 diabetes risk, in addition to confirming the known association with the *TCF7L2* gene. These loci include a non-synonymous polymorphism in the zinc transporter *SLC30A8*, which is expressed exclusively in insulin-producing β -cells, and two linkage disequilibrium blocks that contain genes potentially involved in β -cell development or function (*IDE-KIF11-HHEX* and *EXT2-ALX4*). These associations explain a substantial portion of disease risk and constitute proof of principle for the genome-wide approach to the elucidation of complex genetic traits.

The rapidly increasing prevalence of type 2 diabetes mellitus (T2DM) is thought to be due to environmental factors, such as increased availability of food and decreased opportunity and motivation for physical activity, acting on genetically susceptible individuals. The heritability of T2DM is one of the best established among common diseases and, consequently, genetic risk factors for T2DM have been the subject of intense research¹. Although the genetic causes of many monogenic forms of diabetes (maturity onset diabetes in the young, neonatal mitochondrial and other syndromic types of diabetes mellitus) have been elucidated, few variants leading to common T2DM have been clearly identified and individually confer only a small risk (odds ratio \approx 1.1–1.25) of developing T2DM¹. Linkage studies have reported many T2DM-linked chromosomal regions and have identified putative, causative genetic variants in *CAPN10* (ref. 2), *ENPP1* (ref. 3), *HNF4A* (refs 4, 5) and *ACDC* (also called *ADIPOQ*)⁶. In parallel, candidate-gene studies have reported many T2DM-associated loci, with coding variants in the nuclear receptor *PPARG* (P12A)⁷ and the potassium channel *KCNJ11* (E23K)⁸ being among the very few that have been convincingly replicated. The strongest known (odds ratio \approx 1.7) T2DM association⁹ was recently mapped to the transcription factor *TCF7L2* and has been consistently replicated in multiple populations^{10–20}.

Subjects and study design

The recent availability of high-density genotyping arrays, which combine the power of association studies with the systematic nature of a genome-wide search, led us to undertake a two-stage, genome-wide association study to identify additional T2DM susceptibility loci (Supplementary Fig. 1). In the first stage of this study, we obtained

genotypes for 392,935 single-nucleotide polymorphisms (SNPs) in 1,363 T2DM cases and controls (Supplementary Table 1). In order to enrich for risk alleles²¹, the diabetic subjects studied in stage 1 were selected to have at least one affected first degree relative and age at onset under 45 yr (excluding patients with maturity onset diabetes in the young). Furthermore, in order to decrease phenotypic heterogeneity and to enrich for variants determining insulin resistance and β -cell dysfunction through mechanisms other than severe obesity, we initially studied diabetic patients with a body mass index (BMI) $<30 \text{ kg m}^{-2}$. Control subjects were selected to have fasting blood glucose $<5.7 \text{ mmol l}^{-1}$ in DESIR, a large prospective cohort for the study of insulin resistance in French subjects²².

Genotypes for each study subject were obtained using two platforms: Illumina Infinium Human1 BeadArrays, which assay 109,365 SNPs chosen using a gene-centred design; and Human Hap300 BeadArrays, which assay 317,503 SNPs chosen to tag haplotype blocks identified by the Phase I HapMap²³. Of the 409,927 markers that passed quality control (Supplementary Tables 2 and 3), genotypes were obtained for an average of 99.2% (Human1) and 99.4% (Hap300) of markers for each subject with a reproducibility of $>99.9\%$ (both platforms). Forty-three subjects were removed from analysis because of evidence of intercontinental admixture (Supplementary Fig. 3) and an additional four because their genotype-determined gender disagreed with clinical records. In total, T2DM association was tested for 100,764 (Human1) and 309,163 (Hap300) SNPs representing 392,935 unique loci (Fig. 1). Because of unequal male/female ratios in our cases and controls, we analysed the 12,666 sex-chromosome SNPs separately for each gender.

¹Departments of Human Genetics, ²Medicine and ³Pediatrics, Faculty of Medicine, McGill University, Montreal H3H 1P3, Canada. ⁴McGill University and Genome Quebec Innovation Centre, Montreal H3A 1A4, Canada. ⁵CNRS 8090-Institute of Biology, Pasteur Institute, Lille 59019 Cedex, France. ⁶Endocrinology and Diabetology, University Hospital, Poitiers 86021 Cedex, France. ⁷INSERM U780-IFR69, Villejuif 94807, France. ⁸Endocrinology-Diabetology Unit, Corbeil-Essonnes Hospital, Corbeil-Essonnes 91100, France. ⁹Ontario Institute for Cancer Research, Toronto M5G 1L7, Canada. ¹⁰Montreal Diabetes Research Center, Montreal H2L 4M1, Canada. ¹¹Molecular Nutrition Unit and the Department of Nutrition, University of Montreal and the Centre Hospitalier de l'Université de Montréal, Montreal H3C 3J7, Canada. ¹²Polypeptide Hormone Laboratory and Department of Anatomy and Cell Biology, Montreal H3A 2B2, Canada. ¹³Department of Epidemiology & Public Health, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK. ¹⁴Section of Genomic Medicine, Imperial College London W12 0NN, and Hammersmith Hospital, Du Cane Road, London W12 0HS, UK.

*These authors contributed equally to this work.

Analysis

Markers were selected for assessment in a second cohort using significance thresholds on the basis of the divergence between the observed and expected P -values (Supplementary Figs 4 and 5). These included 28 autosomal SNPs from the Human1 chip ($P < 1 \times 10^{-4}$; Supplementary Table 4) and 43 autosomal SNPs from the Hap300 chip ($P < 5 \times 10^{-5}$; Supplementary Table 5) for a total of 66 unique SNPs representing 44 unique loci. No X-linked marker attained significance, a result that may be due to reduced power of the gender-specific analysis. P -values calculated using 10,000 permutations of the disease state labels identified the same significant associations (Supplementary Tables 4 and 5). Our stage 1 results included the known T2DM association⁹ with the *TCF7L2* SNP rs7903146 ($P = 3.2 \times 10^{-17}$). Several other SNPs at that locus also attained genome-wide significance after correcting for 392,935 tests. In contrast, none of the other previously identified T2DM genes did so (for example, *PPARG*), which is not surprising because our stage 1 had limited power to detect their modest effect and also the arrays did not include the best-associated variants at these loci. However, SNPs tagging four out of seven of these loci attained significance at $P < 0.05$ (Supplementary Table 6).

Because one of the loci showing the strongest T2DM association (rs932206) maps 200 kilobases (kb) telomeric to the lactase gene on 2q21, a region displaying recent positive selection and a north-to-south minor allele frequency (MAF) gradient in Europe, we suspected a spurious association due to population stratification²⁴. This was tested with principal component analysis using 20,323 markers with $\text{MAF} \geq 0.1$, perfectly genotyped in all samples on the Human Hap300 chip, showing no T2DM association in stage 1 ($P > 0.01$) and separated by at least 100 kb. Using the first principal component as a covariate for ancestry differences between cases and controls, we tested for association between rs932206 and disease status. Our result suggests that this apparent association is largely

attributable to ancestry differences ($P = 0.0016$ after adjusting for stratification). However, the selection responsible for the European gradient may be related to metabolic fitness and T2DM risk, and therefore the observed association may not be spurious. Similar testing of the other significant loci did not reveal evidence of stratification and the correction did not affect the statistical significance of their association with T2DM.

We thus prioritized 59 SNPs showing significant association in stage 1, including one of the eight significant *TCF7L2* markers, for rapid confirmation on a larger cohort, using the Sequenom iPLEX assay (Supplementary Fig. 1). We successfully obtained genotypes from 2,617 T2DM cases and 2,894 controls for 57 SNPs (see Supplementary Information). Unlike the stage 1 sample, the affected individuals used in stage 2 were not required to have a family history of T2DM or to be lean (however, severely obese subjects were excluded by requiring $\text{BMI} < 35 \text{ kg m}^{-2}$). We also relaxed the inclusion criteria for control subjects to include individuals with normal fasting glucose levels according to 1997 American Diabetes Association (ADA) criteria ($< 6.1 \text{ mM}$). The SNPs selected for rapid validation were analysed analogously to stage 1 (Supplementary Table 7). In total, eight SNPs representing five unique loci showed significant association after Bonferroni correction was applied for the 57 SNPs tested, based on P -values calculated using 10,000,000 permutations of the disease state labels (Table 1). Because the validation stage samples were selected on the basis of more relaxed inclusion criteria than the stage 1 samples, we used a logistic regression model to investigate the effect of phenotypic variables on T2DM association. The Wald test was used to assess effects of age, sex and BMI on the association between marker and disease, as it is asymptotically equivalent to the Armitage trend test used to detect association in stages 1 and 2. None of the associations (Supplementary Table 7) was substantially changed by considering the effects of these covariates.

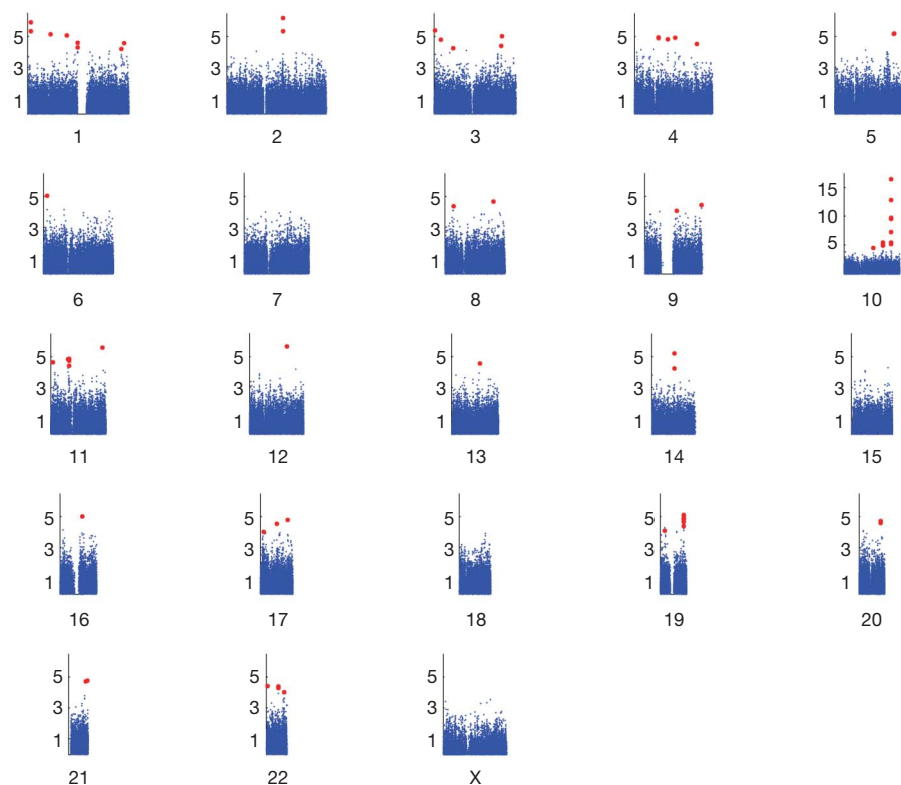


Figure 1 | Graphical summary of stage 1 association results. T2DM association was determined for SNPs on the Human1 and Hap300 chips. The x axis represents the chromosome position from pter; the y axis shows

$-\log_{10}[\text{pMAX}]$, the P -value obtained by the MAX statistic, for each SNP (Note the different scale on the y axis of the chromosome 10 plot.). SNPs that passed the cutoff for a fast-tracked second stage are highlighted in red.

Table 1 | Confirmed association results

SNP	Chromosome	Position (nucleotides)	Risk allele	Major allele	MAF (case)	MAF (ctrl)	Odds ratio (het)	Odds ratio (hom)	PAR	λ_s	Stage 2 pMAX	Stage 2 pMAX (perm)	Stage 1 pMAX	Stage 1 pMAX (perm)	Nearest gene
rs7903146	10	114,748,339	T	C	0.406	0.293	1.65 ± 0.19	2.77 ± 0.50	0.28	1.0546	1.5 × 10 ⁻³⁴	<1.0 × 10 ⁻⁷	3.2 × 10 ⁻¹⁷	<3.3 × 10 ⁻¹⁰	<i>TCF7L2</i>
rs13266634	8	118,253,964	C	C	0.254	0.301	1.18 ± 0.25	1.53 ± 0.31	0.24	1.0089	6.1 × 10 ⁻⁸	5.0 × 10 ⁻⁷	2.1 × 10 ⁻⁵	1.8 × 10 ⁻⁵	<i>SLC30A8</i>
rs1111875	10	94,452,862	G	G	0.358	0.402	1.19 ± 0.19	1.44 ± 0.24	0.19	1.0069	3.0 × 10 ⁻⁶	7.4 × 10 ⁻⁶	9.1 × 10 ⁻⁶	7.3 × 10 ⁻⁶	<i>HHEX</i>
rs7923837	10	94,471,897	G	G	0.335	0.377	1.22 ± 0.21	1.45 ± 0.25	0.20	1.0065	7.5 × 10 ⁻⁶	2.2 × 10 ⁻⁵	3.4 × 10 ⁻⁶	2.5 × 10 ⁻⁶	<i>HHEX</i>
rs7480010	11	42,203,294	G	A	0.336	0.301	1.14 ± 0.13	1.40 ± 0.25	0.08	1.0041	1.1 × 10 ⁻⁴	2.9 × 10 ⁻⁴	1.5 × 10 ⁻⁵	1.2 × 10 ⁻⁵	<i>LOC387761</i>
rs3740878	11	44,214,378	A	A	0.240	0.272	1.26 ± 0.29	1.46 ± 0.33	0.24	1.0046	1.2 × 10 ⁻⁴	2.8 × 10 ⁻⁴	1.8 × 10 ⁻⁵	1.3 × 10 ⁻⁵	<i>EXT2</i>
rs11037909	11	44,212,190	T	T	0.240	0.271	1.27 ± 0.30	1.47 ± 0.33	0.25	1.0045	1.8 × 10 ⁻⁴	4.5 × 10 ⁻⁴	1.8 × 10 ⁻⁵	1.3 × 10 ⁻⁵	<i>EXT2</i>
rs1113132	11	44,209,979	C	C	0.237	0.267	1.15 ± 0.27	1.36 ± 0.31	0.19	1.0044	3.3 × 10 ⁻⁴	8.1 × 10 ⁻⁴	3.7 × 10 ⁻⁵	2.9 × 10 ⁻⁵	<i>EXT2</i>

Significant T2DM associations were confirmed for eight SNPs in five loci. Allele frequencies, odds ratios (with 95% confidence intervals) and PAR were calculated using only the stage 2 data. Allele frequencies in the controls were very close to those reported for the CEU set (European subjects genotyped in the HapMap project). Induced sibling recurrent risk ratios (λ_s) were estimated using stage 2 genotype counts for the control subjects and assuming a T2DM prevalence of 7% in the French population. hom, homozygous; het, heterozygous; major allele, the allele with the higher frequency in controls; pMAX, P-value of the MAX statistic from the χ^2 distribution; pMAX (perm), P-value of the MAX statistic from the permutation-derived empirical distribution (pMAX and pMAX (perm) are adjusted for variance inflation); risk allele, the allele with higher frequency in cases compared with controls.

Identification of four novel T2DM loci

Our fast-track stage 2 genotyping confirmed the reported association for rs7903146 (*TCF7L2*) on chromosome 10, and in addition identified significant associations for seven SNPs representing four new T2DM loci (Table 1). In all cases, the strongest association for the MAX statistic (see Methods) was obtained with the additive model.

The most significant of these corresponds to rs13266634, a non-synonymous SNP (R325W) in *SLC30A8*, located in a 33-kb linkage disequilibrium block on chromosome 8, containing only the 3' end of this gene (Fig. 2a). *SLC30A8* encodes a zinc transporter expressed solely in the secretory vesicles of β -cells and is thus implicated in the final stages of insulin biosynthesis, which involve co-crystallization

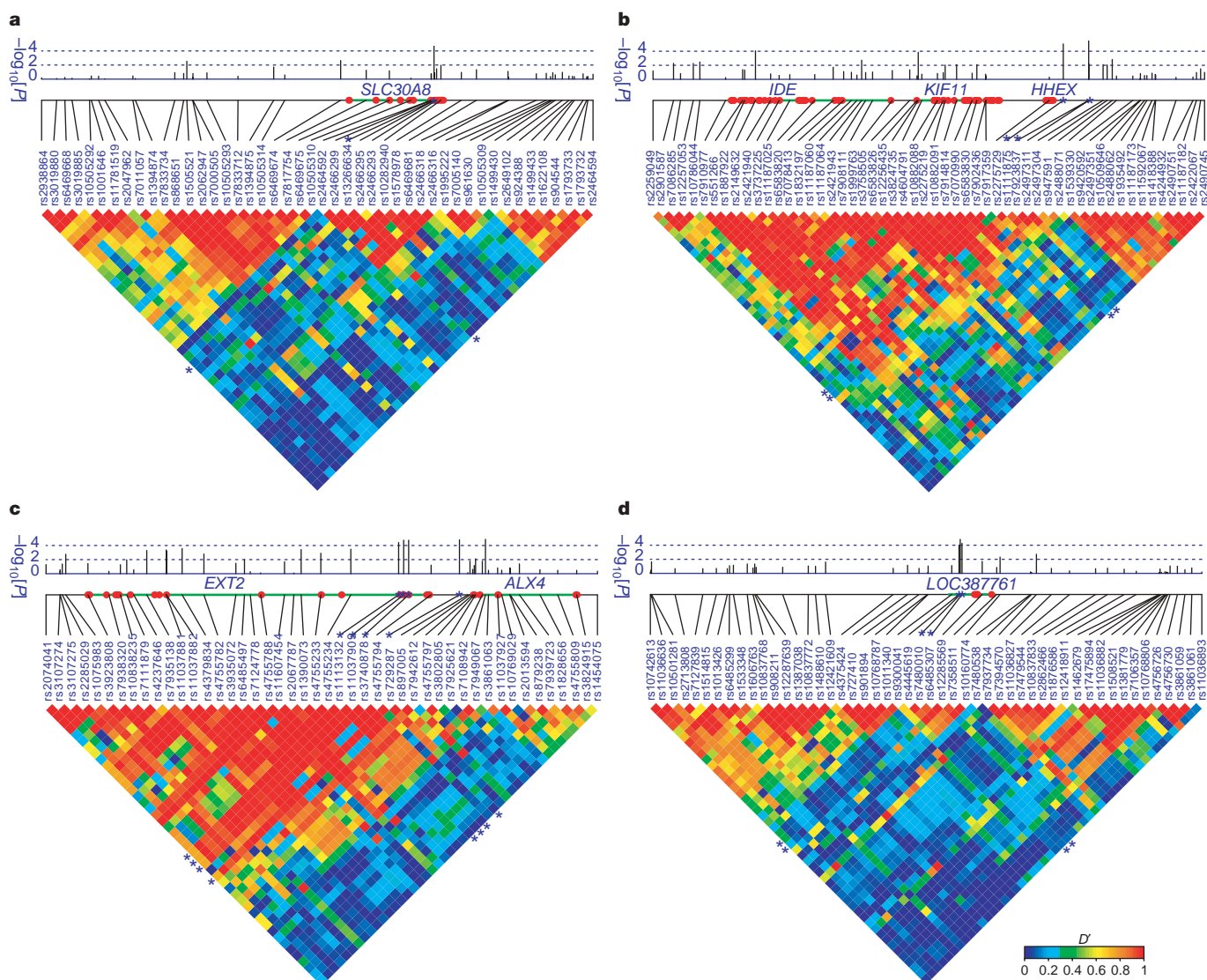


Figure 2 | Pairwise linkage disequilibrium diagrams for four novel T2DM-associated loci. D' was calculated from the stage 1 genotyping data as a fraction of observed linkage disequilibrium over the maximal possible. The bar graph indicates the negative logarithm of the stage 1 P-value for each

SNP. Transcriptional units are indicated by green lines, with exons highlighted in orange. Blue asterisks mark the SNPs chosen for confirmatory studies. **a**, *SLC30A8*; **b**, *IDE-KIF11-HHEX*; **c**, *EXT2-ALX4*; **d**, *LOC387761*.

with zinc. Notably, overexpression of *SLC30A8* in insulinoma cells increases glucose-stimulated insulin secretion²⁵. This finding suggests possible dietary implications and therapeutic approaches with zinc supplementation or, more plausibly, pharmacological manipulation of its transport.

SNPs rs1111875 and rs7923837 are located near the telomeric end of a 270-kb linkage disequilibrium block on chromosome 10 (Fig. 2b), the only one of the novel loci that maps to an interval confirmed in more than one linkage study^{26–29}. The linkage disequilibrium block contains two genes of known biological significance—the insulin-degrading enzyme (*IDE*) and the homeodomain protein *HHEX*—as well as kinesin-interacting factor 11 (*KIF11*). *HHEX* is essential for hepatic and pancreatic development^{30,31} and is a target of the Wnt signalling pathway³², as is *TCF7L2*. Reduction of *IDE* activity by a pharmacological inhibitor increases islet amyloid polypeptide (amylin) accumulation and amylin-mediated cytotoxicity in cultured β -cells³³, whereas *IDE* ablation causes glucose intolerance in knockout mice³⁴. Although *IDE* showed weak T2DM association^{35,36}, these findings were not confirmed in a third well-powered study³⁷. Fine mapping of the *IDE*–*KIF11*–*HHEX* locus in different populations and, ultimately, biological studies will be required to identify the causative variant.

The statistical significance of the top three loci is robust enough to withstand variance-inflation correction for factors of the order of magnitude we observed in stage 1. Two more loci pass Bonferroni correction for 57 markers, but much closer to the cutoff. The first of these involves three SNPs located in introns of exostosin 2 (*EXT2*), at the telomeric end of a 169-kb linkage disequilibrium block on chromosome 11q (Fig. 2c). *EXT2* modulates hedgehog signalling, a pathway involved in early pancreatic development³⁸ and the regulation of insulin synthesis³⁹. This block also contains *ALX4*, a homeodomain protein with possible involvement in the Wnt pathway⁴⁰. Finally, one additional T2DM-associated SNP in chromosome 11 maps to a linkage disequilibrium block that contains the hypothetical gene LOC387761 (Fig. 2d).

To quantify the contribution of these loci to T2DM risk, we calculated the population attributable risk (PAR) for each marker (Table 1). Stepwise logistic regression showed that one SNP per locus explains the entire locus effect and that there was no significant epistasis between loci. Thus, the PAR for the four novel loci together with *TCF7L2* is 70% (see Supplementary Material). In this context, it is worth noting that for three of the four novel loci, the risk allele is the major allele (Table 1). Thus, although our findings can be the source of valuable physiological insights, their contribution to the familial clustering and individual risk prediction of T2DM is relatively small. Of note, for seven of the eight SNPs in Table 1, the risk allele is the ancestral allele, which may be consistent with the hypothesis⁴¹ that the ancestral alleles were adapted to the environment of ancient human populations but today, in a different environment, they increase disease risk. Further population genetics analyses of these loci will allow direct testing of this hypothesis.

Discussion

Our findings permit a number of preliminary insights into the allelic architecture of T2DM susceptibility. We have demonstrated that five relatively common variants (MAF > 0.2) with modest effects (heterozygous relative risk = 1.15–1.65) contribute a significant part towards T2DM risk. Furthermore, we expect that our full stage 2 study will reveal more such loci. Thus, the contribution to T2DM risk of loci with substantial allelic heterogeneity⁴² does not seem to be large. Our results also need to be interpreted in the context of the different selection criteria for the cohorts used in the two stages. We sought to increase effect sizes in stage 1 by excluding obese patients, thus diminishing phenotypic heterogeneity. Consequently, loci conferring risk through effects on insulin secretion or insulin response only in the presence of obesity, if they exist, might not have been detected in our study. In addition, selection for positive family

history probably enriched our stage 1 cohort for individuals carrying risk alleles at a smaller number of loci with stronger effects; this might have compromised our power to detect loci with weak effects. These issues should be addressed in future studies using stratified analysis of larger and more diverse case–control samples.

As might be expected from the clinical characteristics of the stage 1 cohort, the T2DM risk loci we identified seem to involve genes implicated in pancreatic development and the control of insulin secretion. These associations were confirmed in the stage 2 cohort, which is more representative of the general French T2DM population, and potentially highlight the importance of impaired β -cell adaptation to increased metabolic demands in the pathogenesis of T2DM. In addition, these loci may also affect the peripheral response to insulin. For example, *TCF7L2* variants may alter insulin sensitivity in addition to regulating insulin secretion^{11,19}. *HHEX* regulates cell proliferation and tissue specification underlying vascular and hepatic differentiation^{31,43} and *EXT2* is implicated in bone cell proliferation⁴⁴, suggesting a more complex scenario of pleiotropic effects. We anticipate that identification of the causal variants at these genetic loci and their functional consequences will reveal unexpected players in T2DM pathogenesis, and will point to novel mechanisms and targeted therapeutics.

METHODS

Subjects. Detailed characteristics of subjects used in each of the two stages are described in Supplementary Information. Briefly, the stage 1 cases were non-obese (BMI < 30 kg m⁻²) individuals diagnosed with T2DM according to the 1997 criteria of the ADA, who had at least one first degree relative with T2DM. Stage 1 control subjects were selected to have normal fasting plasma glucose and a BMI < 27 kg m⁻². Stage 2 cases required a diagnosis of T2DM by the ADA criteria and BMI < 35 kg m⁻². Stage 2 controls all had normal fasting plasma glucose and BMI < 35 kg m⁻².

Stage 1 whole-genome scan and quality control. Genotyping was performed by labelling 750 ng of genomic DNA and hybridizing it to the Illumina Infinium Human1 and Hap300 BeadArrays, which interrogated 109,365 and 317,503 SNPs, respectively (see Supplementary Information). No significant difference in call rates between cases and controls was seen. Samples successfully genotyped in less than 95% of markers on either array were excluded from analysis, as were subjects whose genotype-inferred gender disagreed with clinical records. Markers were excluded if they deviated significantly from Hardy–Weinberg equilibrium ($P < 0.001$ in the control samples), if they had low MAF (< 0.01 in both the case and control samples), or if they had a call rate < 95% in the case and control samples combined (Supplementary Fig. 2 and Supplementary Tables 2 and 3).

Statistical analysis of stage 1. To identify and correct for possible population stratification, case and control genotypes were analysed using STRUCTURE⁴⁵. For this analysis, our data set was ‘spiked’ with genotypes of unrelated individuals from the four HapMap populations (see Supplementary Information). T2DM association was tested with additive (Armitage trend test), dominant and recessive models for autosomal SNPs⁴⁶, and the largest test statistic obtained from the three models was chosen (MAX statistic). To compensate accurately for testing three models, significance was also estimated against the empirical distribution of the MAX statistic after performing 10,000 permutations of the case and control labels for each marker. To correct for variance inflation owing to systematic genotyping errors or subtle subpopulation structure^{47–50}, the observed χ^2 statistic was adjusted using the Genomic Control method⁴⁸ for each of the three genetic models⁴⁷.

Genotyping and analysis of stage 2. Genotypes for markers selected for fast-track confirmation were obtained using the Sequenom iPLEX assay (Sequenom). Allele detection was performed using matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry. Quality control criteria for markers were the same as in stage 1. Association testing was performed using the MAX statistic and 10,000,000 permutations of the disease state labels. To be considered significant, an association had to involve the same risk allele in both stages. Using the permutation P -values, Bonferroni correction over the 57 SNPs tested gives a significance threshold of $P = 8.8 \times 10^{-4}$.

Received 11 November 2006; accepted 23 January 2007.

Published online 11 February 2007.

1. Permutt, M. A., Wasson, J. & Cox, N. Genetic epidemiology of diabetes. *J. Clin. Invest.* 115, 1431–1439 (2005).

2. Horikawa, Y. *et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genet.* **26**, 163–175 (2000).
3. Meyre, D. *et al.* Variants of *ENPP1* are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nature Genet.* **37**, 863–867 (2005).
4. Love-Gregory, L. D. *et al.* A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 α gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an ashkenazi jewish population. *Diabetes* **53**, 1134–1140 (2004).
5. Silander, K. *et al.* Genetic variation near the hepatocyte nuclear factor-4 α gene predicts susceptibility to type 2 diabetes. *Diabetes* **53**, 1141–1149 (2004).
6. Vasseur, F. *et al.* Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the *APM1* gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. *Hum. Mol. Genet.* **11**, 2607–2614 (2002).
7. Altshuler, D. *et al.* The common PPAR γ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genet.* **26**, 76–80 (2000).
8. Gloyn, A. L. *et al.* Large-scale association studies of variants in genes encoding the pancreatic β -cell K $_{ATP}$ channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) confirm that the *KCNJ11* E23K variant is associated with type 2 diabetes. *Diabetes* **52**, 568–572 (2003).
9. Grant, S. F. *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nature Genet.* **38**, 320–323 (2006).
10. Zhang, C. *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. *Diabetes* **55**, 2645–2648 (2006).
11. Duncott, C. M. *et al.* Polymorphisms in the transcription factor 7-like 2 (*TCF7L2*) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes* **55**, 2654–2659 (2006).
12. Scott, L. J. *et al.* Association of transcription factor 7-like 2 (*TCF7L2*) variants with type 2 diabetes in a Finnish sample. *Diabetes* **55**, 2649–2653 (2006).
13. Groves, C. J. *et al.* Association analysis of 6,736 U.K. subjects provides replication and confirms *TCF7L2* as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* **55**, 2640–2644 (2006).
14. Cauchi, S. *et al.* *TCF7L2* variation predicts hyperglycemia incidence in a French general population: the data from an epidemiological study on the Insulin Resistance Syndrome (DESIR) study. *Diabetes* **55**, 3189–3192 (2006).
15. Chandak, G. R. *et al.* Common variants in the *TCF7L2* gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia* **50**, 63–67 (2007).
16. Florez, J. C. *et al.* *TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N. Engl. J. Med.* **355**, 241–250 (2006).
17. Humphries, S. E. *et al.* Common variants in the *TCF7L2* gene and predisposition to type 2 diabetes in UK European whites, Indian Asians and Afro-Caribbean men and women. *J. Mol. Med.* **84**, 1–10 (2006).
18. Parton, L. E. *et al.* Limited role for SREBP-1c in defective glucose-induced insulin secretion from Zucker diabetic fatty rat islets: a functional and gene profiling analysis. *Am. J. Physiol. Endocrinol. Metab.* **291**, E982–E994 (2006).
19. Saxena, R. *et al.* Common single nucleotide polymorphisms in *TCF7L2* are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* **55**, 2890–2895 (2006).
20. Weedon, M. N. *et al.* Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med.* **3**, e374 (2006).
21. Fingerlin, T. E., Boehnke, M. & Abecasis, G. R. Increasing the power and efficiency of disease-marker case-control association studies through use of allele-sharing information. *Am. J. Hum. Genet.* **74**, 432–443 (2004).
22. Balkau, B. An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome. *Rev. Epidemiol. Sante Publique* **44**, 373–375 (1996).
23. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
24. Campbell, C. D. *et al.* Demonstrating stratification in a European American population. *Nature Genet.* **37**, 868–872 (2005).
25. Chimienti, F. *et al.* *In vivo* expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion. *J. Cell Sci.* **119**, 4199–4206 (2006).
26. Duggirala, R. *et al.* Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am. J. Hum. Genet.* **64**, 1127–1140 (1999).
27. Ghosh, S. *et al.* The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. *Am. J. Hum. Genet.* **67**, 1174–1185 (2000).
28. Meigs, J. B., Panhuysen, C. I., Myers, R. H., Wilson, P. W. & Cupples, L. A. A genome-wide scan for loci linked to plasma levels of glucose and HbA(1c) in a community-based sample of Caucasian pedigrees: The Framingham Offspring Study. *Diabetes* **51**, 833–840 (2002).
29. Wiltshire, S. *et al.* A genomewide scan for loci predisposing to type 2 diabetes in the U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am. J. Hum. Genet.* **69**, 553–569 (2001).
30. Bort, R., Martinez-Barbera, J. P., Beddington, R. S. & Zaret, K. S. *Hex* homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* **131**, 797–806 (2004).
31. Bort, R., Signore, M., Tremblay, K., Martinez Barbera, J. P. & Zaret, K. S. *Hex* homeobox gene controls the transition of the endoderm to a pseudostratified, cell emergent epithelium for liver bud development. *Dev. Biol.* **290**, 44–56 (2006).
32. Foley, A. C. & Mercola, M. Heart induction by Wnt antagonists depends on the homeodomain transcription factor *Hex*. *Genes Dev.* **19**, 387–396 (2005).
33. Bennett, R. G., Hamel, F. G. & Duckworth, W. C. An insulin-degrading enzyme inhibitor decreases amylin degradation, increases amylin-induced cytotoxicity, and increases amyloid formation in insulinoma cell cultures. *Diabetes* **52**, 2315–2320 (2003).
34. Farris, W. *et al.* Insulin-degrading enzyme regulates the levels of insulin, amyloid β -protein, and the β -amyloid precursor protein intracellular domain *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 4162–4167 (2003).
35. Groves, C. J. *et al.* Association and haplotype analysis of the insulin-degrading enzyme (IDE) gene, a strong positional and biological candidate for type 2 diabetes susceptibility. *Diabetes* **52**, 1300–1305 (2003).
36. Karamohamed, S. *et al.* Polymorphisms in the insulin-degrading enzyme gene are associated with type 2 diabetes in men from the NHLBI Framingham Heart Study. *Diabetes* **52**, 1562–1567 (2003).
37. Florez, J. C. *et al.* High-density haplotype structure and association testing of the insulin-degrading enzyme (IDE) gene with type 2 diabetes in 4,206 people. *Diabetes* **55**, 128–135 (2006).
38. Apelqvist, A., Ahlgren, U. & Edlund, H. Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. *Curr. Biol.* **7**, 801–804 (1997).
39. Thomas, M. K., Rastalsky, N., Lee, J. H. & Habener, J. F. Hedgehog signaling regulation of insulin production by pancreatic β -cells. *Diabetes* **49**, 2039–2047 (2000).
40. Boras-Granic, K., Grosschedl, R. & Hamel, P. A. Genetic interaction between *Lef1* and *Alx4* is required for early embryonic development. *Int. J. Dev. Biol.* **50**, 601–610 (2006).
41. Di Rienzo, A. & Hudson, R. R. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet.* **21**, 596–601 (2005).
42. Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69**, 124–137 (2001).
43. Hallaq, H. *et al.* A null mutation of *Hhex* results in abnormal cardiac development, defective vasculogenesis and elevated Vegfa levels. *Development* **131**, 5197–5209 (2004).
44. Stickens, D. *et al.* The *EXT2* multiple exostoses gene defines a family of putative tumour suppressor genes. *Nature Genet.* **14**, 25–32 (1996).
45. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
46. Sasieni, P. D. From genotypes to genes: doubling the sample size. *Biometrics* **53**, 1253–1261 (1997).
47. Clayton, D. G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nature Genet.* **37**, 1243–1246 (2005).
48. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
49. Reich, D. E. & Goldstein, D. B. Detecting association in a case-control study while correcting for population stratification. *Genet. Epidemiol.* **20**, 4–16 (2001

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by Genome Canada, G  n  me Qu  bec, and the Canada Foundation for Innovation. Cohort recruitment was supported by the Association Fran  aise des Diab  tiques, INSERM, CNAMTS, Centre Hospitalier Universitaire Poitiers, La Fondation de France and industrial partners. We thank all individuals who participated as cases or controls in this study. C. Petit, J.-P. Riveline and S. Franc were instrumental in recruitment and S. Brunet, F. Bacot, R. Frechette, V. Catudal, M. Deweirdt, F. Allegaert, P. Laflamme, P. Lepage, W. Astle, M. Leboeuf and S. Leroux provided technical assistance. K. Shazand and N. Foisset provided organizational guidance. Large-scale computations were made possible by the CLUMeq supercomputer facility.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.P. (Constantin.Polychronakos@McGill.ca).

ARTICLES

Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver

Ben Z. Stanger^{1,2,†}, Akemi J. Tanaka¹ & Douglas A. Melton¹

The determinants of vertebrate organ size are poorly understood, but the process is thought to depend heavily on growth factors and other environmental cues. In the blood and central nervous system, for example, organ mass is determined primarily by growth-factor-regulated cell proliferation and apoptosis to achieve a final target size. Here, we report that the size of the mouse pancreas is constrained by an intrinsic programme established early in development, one that is essentially not subject to growth compensation. Specifically, final pancreas size is limited by the size of the progenitor cell pool that is set aside in the developing pancreatic bud. By contrast, the size of the liver is not constrained by reductions in the progenitor cell pool. These findings show that progenitor cell number, independently of regulation by growth factors, can be a key determinant of organ size.

During development, and regeneration, the final shape and size of organs can be restored after cellular loss. This compensation, sometimes called regulative development, is a well-established feature of embryogenesis¹. The regulation that achieves and maintains organ size is accomplished, in many cases, by systemic or 'extrinsic' factors². For example, overall size in organisms ranging from *Drosophila* to *Homo sapiens* depends on nutritional status and endocrine factors including growth hormone, insulin and insulin-like growth factor^{3–7}. Extrinsic signals similarly regulate size in many vertebrate tissues, including blood, liver, muscle and the central nervous system. Such signals act by controlling cell proliferation or by modulating cell death³.

In a few cases, however, it seems that final organ size depends on 'intrinsic' factors. This has been recognized in vertebrates through transplantation experiments in which a grafted tissue grows to a 'pre-determined' size regardless of the host environment. For example, transplantation of limb buds from the salamander and chick demonstrated that the limb grows to a size dictated by the donor limb bud, rather than the host^{8,9}. Furthermore, in the blood—a tissue whose overall size in the adult is highly dependent on regulatory influences—both intrinsic and extrinsic factors seem to be involved in determining haematopoietic stem cell (HSC) number during development^{10,11}. Although these experiments point to the possible existence of autonomous size programmes, it is unclear what mechanisms underlie intrinsic control.

We have examined the issue of mammalian organ size regulation in the pancreas and liver with a new approach. We used two different methods—cell ablation and tissue complementation—to perturb precursor cell number during the earliest stages of pancreatic and liver development. These manipulations have profound effects on the final size of the pancreas, reflecting an inability of pancreatic progenitor cells to compensate significantly in response to a reduction in number. The liver, by contrast, exhibits robust and precise compensatory growth in response to cell ablation. Overall, these results suggest that the size of some organs (including the pancreas) may be

relatively fixed early in development and that growth factors may have a limited or more permissive role in determining final size.

Conditional progenitor cell ablation

The epithelial lineages of the pancreas (exocrine, endocrine and ductal) arise from cells that express the transcription factor *Pdx1* (pancreatic and duodenal homeobox 1) during development¹²; after birth, *Pdx1* expression becomes largely confined to insulin-producing β -cells. We devised two methods to perturb the number of *Pdx1*-expressing (*Pdx1*⁺) cells in development. First, the *Pdx1*^{tTA} strain, in which the tetracycline transactivator (tTA) coding sequence is integrated in the endogenous *Pdx1* locus¹³, was crossed to *tetO*^{DTA/DTA} mice, in which expression of the diphtheria toxin A chain (DTA) is controlled by upstream tetracycline operator sequences¹⁴. This cross yielded *Pdx1*^{+/+}; *tetO*^{DTA/+} (+/DTA) and *Pdx1*^{tTA/+}; *tetO*^{DTA/+} (tTA/DTA) embryos, in which the elimination of *Pdx1*⁺ progenitors could be regulated by altering the dose/timing of tetracycline addition.

Pregnant mice were given tetracycline to repress the transcriptional activity of tTA¹⁵ or left untreated to activate tTA-mediated transcription of DTA (Fig. 1a), and offspring were examined at birth (E18.5). Although body weight was maintained regardless of genotype or tetracycline treatment regimen (data not shown), tTA/DTA newborns that had not received tetracycline were virtually apapneatic (Figs 1b and 2e; see also Supplementary Fig. 1). By contrast, pancreata from tTA/DTA embryos exposed to tetracycline throughout pregnancy exhibited a normal appearance and preserved weight (Figs 1b and 2a; see also Supplementary Fig. 1).

The consequences of transient progenitor cell ablation were examined by inducing, and then repressing, DTA expression (Fig. 2). Tetracycline was withheld from pregnant females until mid-gestation and subsequently added at specific time points to halt the ablation of progenitor cells. Pancreata from tTA/DTA embryos in which ablation was halted at E9.5 appeared grossly normal at E18.5 and exhibited a small reduction in weight compared to control pancreata

¹Department of Molecular and Cellular Biology, Harvard Stem Cell Institute, and Howard Hughes Medical Institute, Cambridge, Massachusetts 02138, USA. ²Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [†]Present address: Division of Gastroenterology and Abramson Family Cancer Research Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

(Fig. 2b; see also Supplementary Fig. 1). By contrast, embryos in which progenitor cell ablation was halted starting on, or after, E11.5 had a nearly complete absence of the pancreas (Fig. 2d; see also Supplementary Fig. 1). This latter result demonstrates that if $Pdx1^{+}$ cells are efficiently eliminated before \sim E12–E12.5, formation of $Pdx1^{+}$ pancreatic progenitor cells with the ability to rescue pancreatic development does not occur. Stated otherwise, these results suggest that all of the $Pdx1^{+}$ progenitors needed to make the pancreas are generated during the embryonic period spanning approximately E8.5 to E12.5.

Direct evidence that DTA expression reduces the pancreatic progenitor pool was obtained by staining sectioned pancreatic buds with a $Pdx1$ antibody. $Pdx1^{+}$ precursors are first detected at approximately E8.5–9.0 (ref. 16), so the absence of tetracycline up to and including E9.5 is expected to have a minor effect. Indeed, embryos subjected to ablation through E9.5 exhibited a mild reduction (8%) in the size of the $Pdx1^{+}$ progenitor pool, whereas a more marked reduction (61%) was observed after ablation through E10.5 (Supplementary Fig. 2a, c). Ablation through later time points resulted in the loss of pancreatic buds and the near absence of pancreatic progenitors (Supplementary Fig. 2a, c). These results show that death of pancreatic progenitor cells, and resulting loss of pancreatic tissue mass, can be regulated by tetracycline in these mice.

Lack of significant growth compensation

To assess the capacity for compensatory growth, we generated embryos in which many, but not all, progenitor cells were ablated. On the basis of the results above, progenitor ablation up to and including E10.5 was predicted to provide an appropriate experimental condition for testing the ability of pancreas size to normalize. Indeed, ablation of progenitor cells in tTA/DTA embryos through E10.5, followed by a 2-day recovery, resulted in pancreatic buds that were significantly smaller than controls (Supplementary Fig. 2b).

Under these conditions, a reproducible 60–80% reduction in newborn pancreas size was observed (Fig. 2c; see also Supplementary Fig. 1). This result is surprising because it shows that, after a reduction in the $Pdx1^{+}$ progenitor pool at the early bud stage, the remaining pancreatic progenitors are able to grow, divide and differentiate, but cannot increase the rate or number of cell divisions to make an organ of normal size. The DTA-mediated death of a fraction of

$Pdx1^{+}$ progenitor cells, coupled with an inability of the remaining cells to compensate for this reduced cell number, causes the pancreas to be small at birth. The distal stomach and proximal intestine, which normally express $Pdx1$ during development and which therefore might be expected to show some loss, were not noticeably affected in tTA/DTA embryos under any ablation conditions; this may be a result of silencing of the $tetO^{DTA}$ locus in these tissues (data not shown).

One potential explanation for a small pancreas after transient progenitor ablation is the possibility that a subset of progenitor cells was selectively eliminated. Because exocrine cells constitute over 80% of tissue mass, loss of exocrine progenitor cells might account for the reduced size¹². To address this possibility, we determined the relative area occupied by exocrine, ductal and endocrine cells from tTA/DTA pancreata in which progenitor ablation had been halted at E10.5. These pancreata contained all lineages, although the area occupied by insulin-positive β -cells was decreased approximately twofold compared to +/DTA littermates (Supplementary Fig. 3). Because β -cells make up a small percentage (<5%) of overall pancreatic mass, faulty specification of endocrine progenitor cells does not account for the reduction in pancreas size.

Another possibility is that compensatory growth occurs, but does so very slowly. Specifically, it might take longer than the 8 days between DTA repression at E10.5 and analysis at E18.5 for size normalization to occur. To address this, we followed cohorts of mice (in which progenitor cell ablation had been halted at E10.5) for approximately 1, 3, 4, or 11 weeks after birth. Similar to their newborn counterparts, the pancreata of these mice remained proportionately small, exhibiting a growth rate that matched that of littermate controls (Fig. 3). Mice in these cohorts displayed marked glucose intolerance, possibly reflecting the relative and absolute decrease in β -cells noted above. Thus, a reduction in the number of progenitor cells during an early and brief window of pancreatic development has a lasting impact on pancreas size in adult animals, persisting for 3 months after the original developmental deficit.

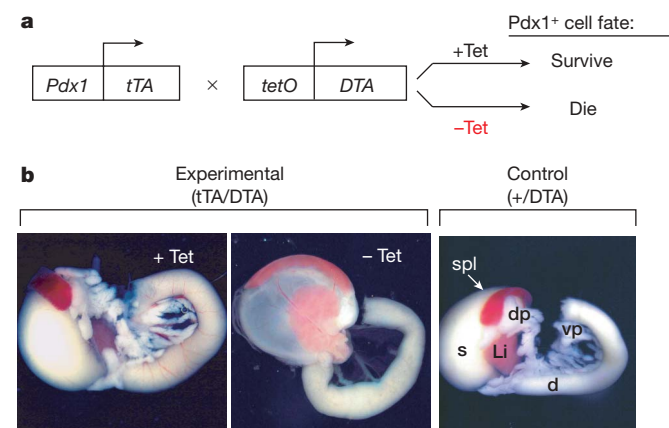


Figure 1 | Pancreatic progenitor cell ablation. **a**, Experimental design. Two classes of embryos are generated by crossing $Pdx1^{tTA/+}$ and $tetO^{DTA/+}$ mice: $Pdx1^{+/+}; tetO^{DTA/+}$ (Control; +/DTA) and $Pdx1^{tTA/+}; tetO^{DTA/+}$ (Experimental; tTA/DTA). In untreated tTA/DTA embryos (–Tet), $Pdx1^{+}$ progenitors are killed after tTA-mediated expression of DTA. Treatment with tetracycline (+Tet) represses tTA-mediated transcription, allowing survival of $Pdx1^{+}$ progenitors. **b**, Pregnant females were given plain water or water containing tetracycline throughout gestation; embryos were examined at E18.5. The midgut of a control (+/DTA) embryo is shown (right panel) to illustrate normal structures. d, duodenum; dp, dorsal pancreas; Li, liver; s, stomach; spl, spleen; vp, ventral pancreas.

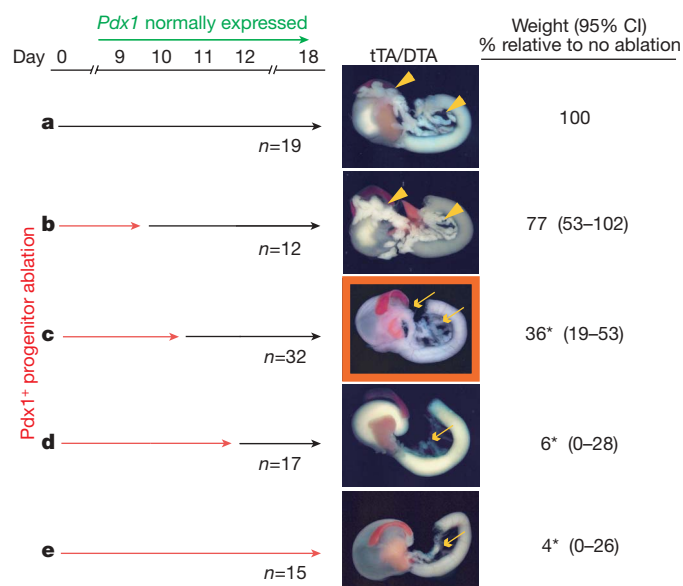


Figure 2 | Lack of compensatory growth during pancreas development. **a–e**, Pregnant mice carrying tTA/DTA embryos were given tetracycline either throughout embryogenesis (**a**) or at various time points during early pancreas organogenesis (**b**, E9.5; **c**, E10.5; **d**, E11.5; **e**, no tetracycline) and embryos were examined at E18.5. Black arrows denote the period of tetracycline administration (DTA repressed); red arrows denote the period during which tetracycline was omitted, resulting in ablation. Weights were calculated as a percentage of non-ablated control (**a**) and are displayed along with 95% confidence intervals (CI) and the number of tTA/DTA embryos examined. Asterisk, $P < 0.001$.

Complementing the pancreatic epithelium

These observations support the notion that pancreatic progenitor cell growth follows an intrinsic or autonomous programme, one that does not aim to achieve a certain final size for the whole organ. However, it is conceivable that other factors contributed to the failure of a limited number of progenitor cells to compensate. For example, dead cells generated by DTA expression might have inhibited expansion of remaining $Pdx1^{+}$ progenitors, or a subset of progenitor cells with a greater capacity for compensatory growth may have been selectively killed by the DTA ablation scheme. To address these possibilities, we developed a system wherein progenitor number could be approached from the opposite direction, where we added back progenitors to dictate the starting number of cells in the pancreatic anlage.

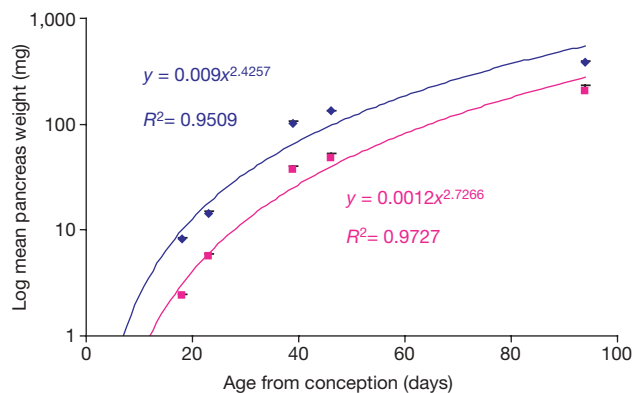
Because homozygous inactivation of $Pdx1$ prevents pancreas development^{17,18}, we hypothesized that a chimaeric embryo derived by combining $Pdx1$ -deficient blastocysts with wild-type embryonic stem (ES) cells would exhibit selective growth and maturation of ES cell progeny during pancreatic development. Furthermore, if the $Pdx1$ mutation acts cell-autonomously—as suggested by *in vitro* recombination experiments¹⁹—the entire pancreatic epithelium generated by such a complementation scheme would be derived from the ES cells. A similar approach has been used successfully in mice to create an ES-cell-derived lymphoid system²⁰ and lens²¹.

We used $Pdx1^{lacZ}$ mice, which harbour a β -galactosidase cDNA in place of $Pdx1$ coding sequences¹⁸. Wild-type ES cells were injected into blastocysts generated by an intercross of $Pdx1^{lacZ/+}$ heterozygous mice, and embryos were collected at later stages (Fig. 4a). As predicted, pancreata from $Pdx1^{+/+}$ and $Pdx1^{lacZ/+}$ embryos appeared normal. Notably, several chimaeric embryos derived from $Pdx1^{lacZ/lacZ}$ blastocysts also appeared to have a normal pancreas. The structures that express $Pdx1$ —stomach, pancreas and duodenum—stain blue in $Pdx1^{lacZ/+}$ embryos (Fig. 4c), allowing us to distinguish between blastocyst-derived and ES-derived epithelial cells. In all cases, the pancreas arising from a rescued $Pdx1^{lacZ/lacZ}$ blastocyst lacked β -galactosidase activity (Fig. 4e), indicating that it was formed from

donor ES cells. $Pdx1^{lacZ/lacZ}$ embryos without ES cell contribution had a blue 'dorsal ductule', representing the arrested product of a dorsal pancreatic bud as previously noted¹⁸, but no pancreas (Fig. 4d). When ES cells tagged with yellow fluorescent protein (YFP)²² were used to make chimaeras, the entire pancreatic epithelium of complemented $Pdx1$ -deficient chimaeras was YFP⁺ (Fig. 4e, inset; see also Supplementary Fig. 4). These results demonstrate unambiguously the cell-autonomous nature of $Pdx1$ function and the ability of wild-type ES cells to give rise to a normal pancreas when introduced into a $Pdx1$ mutant blastocyst.

Size variation in complemented pancreata

Chimaeras generated by blastocyst injection exhibit a wide range of ES-cell contributions. Although the factors that determine the extent of chimaerism are poorly understood, it seems that only one or two ES cells contribute to each chimaeric embryo²³. Because of this inherent variation, the tissue complementation system allowed us to ask whether a limited number of competent cells can undergo compensatory growth. If a small number of progenitor cells are able to compensate, we would expect to observe a normal-sized pancreas in all cases, regardless of the degree of chimaerism. Conversely, a lack of



Number analysed:

	Birth	1 wk	3 wk	4 wk	11 wk
+ /DTA	29	12	25	12	19
tTA/DTA	32	6	17	7	12

Figure 3 | Small pancreata do not exhibit catch-up growth. To generate adult mice with small pancreata, pregnant females carrying $Pdx1^{tTA/+}; tetO^{DTA/+}$ (tTA/DTA; pink) and $Pdx1^{+/+}; tetO^{DTA/+}$ (+/DTA; blue) embryos were given tetracycline at E10.5 to halt ablation. Pancreas weights were measured at birth and approximately 1, 3, 4, or 11 weeks of age. Mean (\pm s.e.m.) pancreas weights of experimental and control littermates were plotted on a log₁₀ scale against age, and best-fit curves were generated. Experimental pancreata grow at approximately the same rate as control pancreata but remain proportionately small. The number of animals analysed at each experimental time point is shown below.

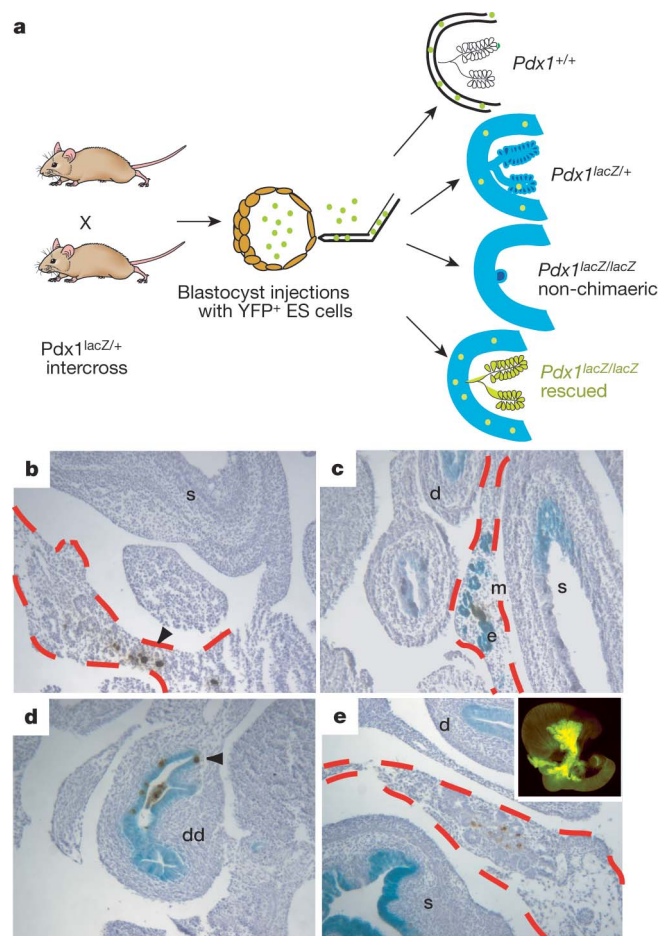


Figure 4 | $Pdx1$ -deficient blastocyst complementation. **a**, Strategy. YFP-tagged ES cells are injected into blastocysts generated from an intercross of $Pdx1^{lacZ/+}$ mice. Homozygous $Pdx1^{lacZ/lacZ}$ embryos exhibiting pancreas complementation should express YFP but not β -galactosidase. **b–e**, Glucagon (brown) and β -galactosidase (blue) staining of E14.5 midguts; pancreata are outlined in red. **b**, Chimaeric $Pdx1^{+/+}$ embryos exhibit pancreatic glucagon staining (arrowhead) but no β -galactosidase activity. **c**, **d**, Glucagon staining and β -galactosidase activity are both present in heterozygous $Pdx1^{lacZ/+}$ pancreata (**c**) and in the dorsal ductules¹⁸ of $Pdx1^{lacZ/lacZ}$ embryos (**d**). **e**, Complemented $Pdx1^{lacZ/lacZ}$ pancreata lack β -galactosidase activity but are YFP⁺ (inset). d, duodenum; dd, dorsal ductule; e, epithelium; m, mesenchyme; s, stomach.

compensatory growth would result in variation in pancreas size (Fig. 5a).

Although the vast majority of chimaeras generated from $Pdx1^{-/-}$ blastocysts contained some pancreatic tissue (23 out of 27; 85%), significant variation in pancreas size was observed (Fig. 5b–d). In each case, the pancreas, when formed, had exocrine, endocrine and duct cells, but the overall size of the organ varied considerably, ranging from normal to microscopic (Supplementary Table 1). Furthermore, the extent of chimaerism in complemented embryos, measured by YFP fluorescence in extra-pancreatic tissues, was correlated with the size of the pancreas (Fig. 5e–g; see also Supplementary Fig. 5). These results are consistent with the hypothesis that a limited number of competent pancreatic progenitor cells come to populate the developing endodermal gut tube and that the fates of these ES-derived cells are fixed at a certain time. Subsequently, these cells and their progeny are unable to undergo compensatory growth to produce a pancreas of normal size.

Growth compensation in the liver

We wished to determine whether other tissues develop with similar constraints on growth. The liver was chosen for comparison because of its close developmental relationship to the pancreas. Both organs are derived from adjacent regions of endoderm and there is evidence for a bi-potential progenitor cell with the capacity

to give rise to either liver or pancreas²⁴. Furthermore, the adult liver undergoes robust regeneration after injury or resection²⁵, in contrast to the pancreas, which has a more limited capacity for regeneration.

To ablate hepatic progenitor cells, we used a transgenic strain in which the promoter for LAP (liver-enriched transcriptional activator protein), a marker of early hepatic progenitor cells²⁶, drives tTA-mediated gene expression in the liver at a level at least 200-fold higher than that of any other tissue^{15,27}. Using this strain, the elimination of LAP⁺ hepatic progenitors could be regulated by mating LAP^{tTA} mice to $tetO^{DTA}$ mice and altering the dose/timing of tetracycline addition. Ablation through E13.5 resulted in a reduction in the hepatic progenitor pool of at least 65% (Supplementary Figs 6 and 7), enabling us to perform an experiment comparable to the transient ablation of pancreatic progenitors described above. After an initial pulse of ablation²⁸, DTA expression was repressed with tetracycline, and liver size was determined 4 days later. Under these conditions, and distinct from the corresponding result in the pancreas, the livers of LAP-tTA/DTA embryos appeared normal (compare orange boxed panels in Figs 6 and 2c). This ‘catch-up’ growth occurred rapidly and with great precision (Fig. 6b; see also Supplementary Fig. 7). Thus, in contrast to the inability of the pancreas to normalize size long after a loss of early progenitor cells, the liver regulates growth to achieve a normal size after only 4 days (compare Figs 6b and 3).

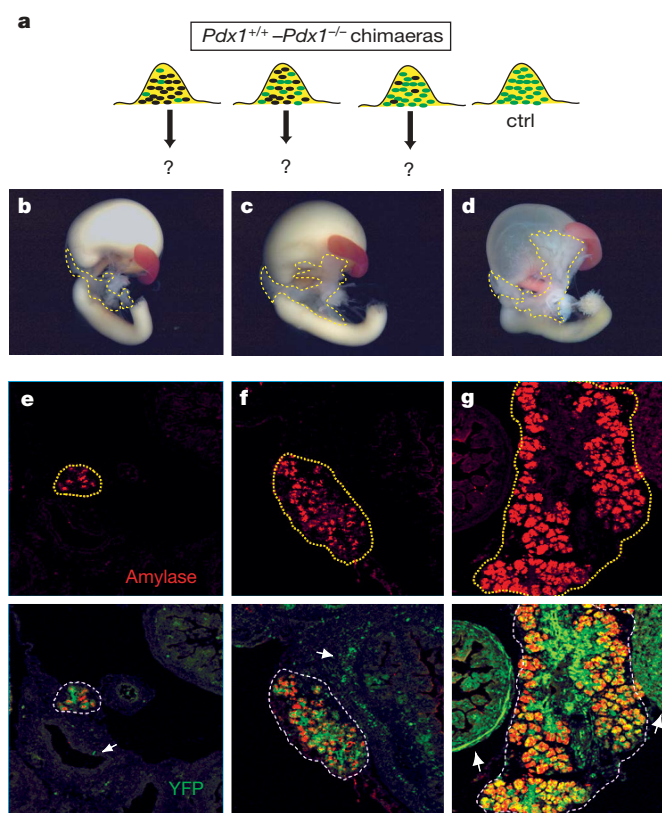


Figure 5 | Size variation after pancreas complementation. **a**, Injection of wild-type ES cells into $Pdx1^{-/-}$ blastocysts yields chimaeras with variable ES contribution. The total number of cells within the pancreatic anlage is the same in chimaeric and wild-type (ctrl) mice; however, the relative percentage of competent $Pdx1^{+}$ cells (green) and non-competent $Pdx1^{-}$ cells (black) varies. The ability of a reduced number of competent cells to give rise to a normal pancreas is tested by examining newborn pancreata from multiple chimaeras. **b–d**, Dissected guts from three complemented $Pdx1^{-/-}$ embryos (E18.5), showing variation in pancreas size (outlined). **e–g**, Top row: three complemented pancreata exhibiting size variation (different from those above) were stained with amylase (red), yielding microscopic (**e**), small (**f**) and normal (**g**) pancreata. Bottom row: the degree of chimaerism (green; arrows) correlates with pancreas size.

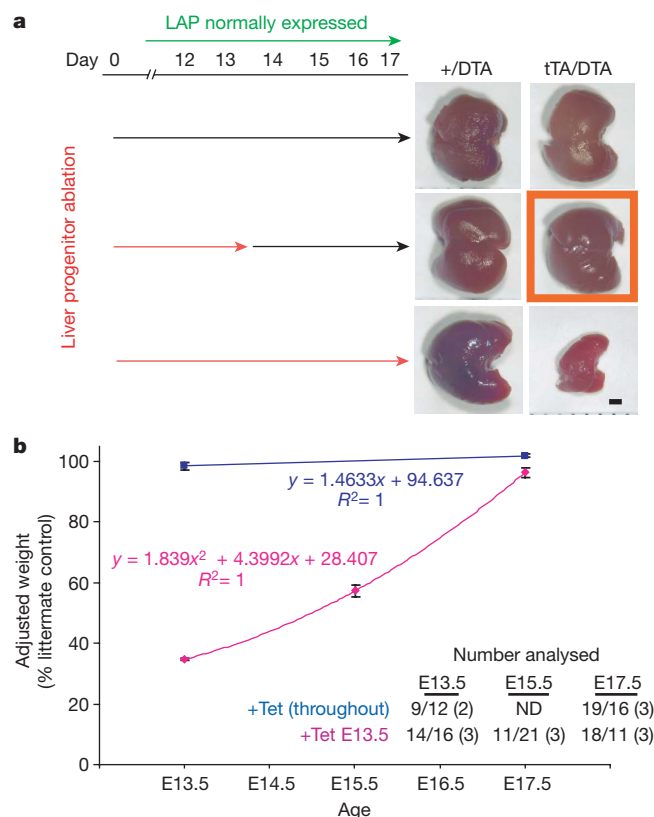


Figure 6 | Compensatory growth during liver development. **a**, Pregnant mice carrying LAP-tTA/DTA and +/DTA embryos were given plain drinking water or water containing tetracycline at designated time points and embryos were examined at E17.5. Black arrows denote the period of tetracycline administration (DTA repressed); red arrows denote the period during which tetracycline was omitted (ablation). Repression of DTA expression from E13.5–E17.5 results in a complete recovery of liver size (boxed panel). Scale bar, 1 mm. **b**, Tetracycline was provided either throughout gestation (blue) or beginning at E13.5 (pink). LAP-tTA/DTA liver weight was plotted as a percentage of +/DTA littermate control weight (mean \pm s.e.m.) using best-fit analysis. The number of embryos examined at each time point (experimental/control) is shown, along with the number of litters examined (parentheses). ND, not determined.

Discussion

Given the highly regulative nature of vertebrate development, we expected that reducing the number of early tissue-specific progenitor cells would not perturb final organ size. This expectation was based on numerous precedents; for example, removing most of the cells in the embryonic limb field in amphibians or chickens has no effect on final limb size because compensatory proliferation makes up for the loss²⁹. Similar compensatory proliferation occurs in *Drosophila* imaginal discs after cell loss, resulting in normal tissue size³⁰. And in the blood and central nervous system, growth factors and apoptotic factors regulate final organ size by controlling cell number^{31,32}. Contrary to these predictions, our studies show that compensatory growth during pancreas development is either quite limited or does not occur at all. Thus, embryonic progenitor cells represent a critical and limiting determinant of pancreas size.

Although the prediction of growth compensation was not borne out in our pancreas studies, it proved to be true for liver size. The adult liver exhibits robust recovery of mass after injury or partial removal. We find that the embryonic liver shares this capacity for compensatory growth and does so with an impressive speed and accuracy. Four days after losing roughly two-thirds of its mass, the embryonic liver exhibits a size that is almost identical to that of controls. By contrast, the pancreas fails to compensate after a similar loss of mass 3 months after the initial insult. Thus, two closely related endoderm derivatives—the pancreas and liver—differ markedly in their response to developmental alterations in progenitor cell number.

We did not perform tissue complementation in the liver, although comparable studies have recently been reported³³. Chimaeric embryos containing a mixture of wild-type and *Hex*-deficient liver progenitor cells exhibit progressive loss of the *Hex*-null cells during liver morphogenesis. In these chimaeras, the remaining wild-type cells increase their proliferation and compensate for this loss, giving rise to a normally sized liver bud³³.

Factors that regulate cellular proliferation are undoubtedly involved in pancreatic growth control. FGF10 is an important mediator of pancreas growth³⁴, acting through Notch signalling to balance proliferation and differentiation^{35,36}. Wnt signalling also has a principal role in pancreas growth during development^{37,38}, and recent observations suggest that stabilization of β -catenin results in an increase in pancreas size through effects on proliferation³⁹. The results reported here suggest that despite a developmental requirement for growth factors, these factors cannot mediate a significant compensatory response in the pancreas.

Our analysis draws attention to the embryonic progenitor cell in setting final organ size. Committed progenitor cells, in contrast to stem cells, are defined by their limited capacity for division and self-renewal. Thus, it is possible that pancreatic progenitor cell proliferation is autonomously restricted and that each progenitor cell is capable of giving rise to only a fixed amount of tissue. In the central nervous system, for example, a 'counting mechanism' limits the number of times a glia precursor cell divides⁴⁰. Pancreas size might be dictated through a similar mechanism, such that the size of the progenitor pool specified during endoderm patterning might dictate final organ size. Conversely, size information might be encoded within the pancreatic mesenchyme, which in turn might control the growth of the epithelium in a manner that is not subject to compensatory growth.

The mechanisms that regulate organ size will probably be understood in the context of two related and general questions: (1) to what extent do different organs develop from self-renewing stem cells versus committed progenitor cells; and (2) to what extent is size determined by cell-autonomous and non-cell-autonomous mechanisms. Although it has been suggested that tissues are generated during fetal life by a hierarchy of stem cell to progenitor cell to progeny cell⁴¹, it is unknown for most organs whether it is stem cells or committed progenitors that constitute the substrate for organogenesis. Our

previous studies of the pancreas suggested that stem cells do not contribute significantly to the maintenance of adult islets⁴². The current study highlights the possibility that this adult state is presaged by an absence of pancreatic stem cells during embryonic development. Moreover, these results raise the question of whether there are two kinds of tissues: some, like the liver, that exhibit regulated growth in development and robust regeneration in adulthood because of a reliance on extrinsic signals; and others, like the pancreas, that exhibit reduced regenerative capacity in adulthood because of intrinsic growth constraints that are imprinted early in development.

METHODS

A detailed description of materials and methods is provided in Supplementary Information.

Mice. Conditional ablation of *Pdx1*⁺ progenitor cells was achieved by mating male *Pdx1*^{TA/+} and female *tetO*^{DTA/DTA} mice and checking for vaginal plugs daily. Tetracycline (1 mg ml⁻¹) was added to the drinking water of pregnant females at specific developmental time points as described in the text. *tetO*^{DTA/DTA} mice were maintained as an inbred strain through brother–sister mating. *Pdx1*^{TA/+} mice were of a mixed genetic background. All strains (*Pdx1*^{TA}, *tetO*^{DTA}, *LAP*^{TA}, *Pdx1*^{lacZ} and *tetO*^{lacZ}) have been described previously^{13–15,18,28}.

Pancreas complementation. *Pdx1*-deficient embryos were generated by intercrossing *Pdx1*^{TA/+} heterozygotes. For early studies (for example, those presented in Fig. 4), only the *Pdx1*^{lacZ} strain was used, whereas later studies used an intercross of *Pdx1*^{lacZ} and *Pdx1*^{TA} mice; both classes of matings yielded identical results. Blastocysts were injected with 10–20 YC5 ES cells, which express YFP after incorporation into a wide variety of tissues²².

Received 15 May; accepted 29 November 2006.

Published online 28 January 2007.

- Gilbert, S. F. *Developmental Biology* (Sinauer, Sunderland, Massachusetts, 2000).
- Bryant, P. J. & Simpson, P. Intrinsic and extrinsic control of growth in developing organs. *Q. Rev. Biol.* **59**, 387–415 (1984).
- Conlon, I. & Raff, M. Size control in animal development. *Cell* **96**, 235–244 (1999).
- Britton, J. S., Lockwood, W. K., Li, L., Cohen, S. M. & Edgar, B. A. *Drosophila*'s insulin/P13-kinase pathway coordinates cellular metabolism with nutritional conditions. *Dev. Cell* **2**, 239–249 (2002).
- Rulifson, E. J., Kim, S. K. & Nusse, R. Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. *Science* **296**, 1118–1120 (2002).
- Ikeya, T., Galic, M., Belawat, P., Nairz, K. & Hafen, E. Nutrient-dependent expression of insulin-like peptides from neuroendocrine cells in the CNS contributes to growth regulation in *Drosophila*. *Curr. Biol.* **12**, 1293–1300 (2002).
- Colombani, J. et al. Antagonistic actions of ecdysone and insulins determine final size in *Drosophila*. *Science* **310**, 667–670 (2005).
- Twitty, V. C. & Schwind, J. L. The growth of eyes and limbs transplanted heteroplastically between two species of *Ambystoma*. *J. Exp. Zool.* **59**, 61–86 (1931).
- Wolpert, L. Cellular basis of skeletal growth during development. *Br. Med. Bull.* **37**, 215–219 (1981).
- Muller-Sieburg, C. E., Cho, R. H., Sieburg, H. B., Kupriyanov, S. & Riblet, R. Genetic control of hematopoietic stem cell frequency in mice is mostly cell autonomous. *Blood* **95**, 2446–2448 (2000).
- Robin, C. et al. An unexpected role for IL-3 in the embryonic development of hematopoietic stem cells. *Dev. Cell* **11**, 171–180 (2006).
- Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3⁺ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447–2457 (2002).
- Holland, A. M., Hale, M. A., Kagami, H., Hammer, R. E. & MacDonald, R. J. Experimental control of pancreatic development and maintenance. *Proc. Natl Acad. Sci. USA* **99**, 12236–12241 (2002).
- Lee, P. et al. Conditional lineage ablation to model human diseases. *Proc. Natl Acad. Sci. USA* **95**, 11371–11376 (1998).
- Kistner, A. et al. Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. *Proc. Natl Acad. Sci. USA* **93**, 10933–10938 (1996).
- Guz, Y. et al. Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. *Development* **121**, 11–18 (1995).
- Jonsson, J., Carlsson, L., Edlund, T. & Edlund, H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606–609 (1994).
- Offield, M. F. et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983–995 (1996).
- Ahlgren, U., Jonsson, J. & Edlund, H. The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. *Development* **122**, 1409–1416 (1996).

20. Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proc. Natl Acad. Sci. USA* **90**, 4528–4532 (1993).
21. Liegeois, N. J., Horner, J. W. & DePinho, R. A. Lens complementation system for the genetic analysis of growth, differentiation, and apoptosis *in vivo*. *Proc. Natl Acad. Sci. USA* **93**, 1303–1307 (1996).
22. Hadjantonakis, A. K., Macmaster, S. & Nagy, A. Embryonic stem cells and mice expressing different GFP variants for multiple non-invasive reporter usage within a single animal. *BMC Biotechnol.* **2**, 11 (2002).
23. Wang, Z. & Jaenisch, R. At most three ES cells contribute to the somatic lineages of chimeric mice and of mice produced by ES-tetraploid complementation. *Dev. Biol.* **275**, 192–201 (2004).
24. Deutsch, G., Jung, J., Zheng, M., Lora, J. & Zaret, K. S. A bipotential precursor population for pancreas and liver within the embryonic endoderm. *Development* **128**, 871–881 (2001).
25. Taub, R. Liver regeneration: from myth to mechanism. *Nature Rev. Mol. Cell Biol.* **5**, 836–847 (2004).
26. Westmacott, A., Burke, Z. D., Oliver, G., Slack, J. M. & Tosh, D. C/EBP α and C/EBP β are markers of early liver development. *Int. J. Dev. Biol.* **50**, 653–657 (2006).
27. Lavon, I. *et al.* High susceptibility to bacterial infection, but no liver dysfunction, in mice compromised for hepatocyte NF- κ B activation. *Nature Med.* **6**, 573–577 (2000).
28. Furth, P. A. *et al.* Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. *Proc. Natl Acad. Sci. USA* **91**, 9302–9306 (1994).
29. Kalthoff, K. *Analysis of Biological Development* (McGraw-Hill, New York, 1996).
30. Gallant, P. Myc, cell competition, and compensatory proliferation. *Cancer Res.* **65**, 6485–6487 (2005).
31. Hidalgo, A. & French-Constant, C. The control of cell number during central nervous system development in flies and mice. *Mech. Dev.* **120**, 1311–1325 (2003).
32. Morrison, S. J., Uchida, N. & Weissman, I. L. The biology of hematopoietic stem cells. *Annu. Rev. Cell Dev. Biol.* **11**, 35–71 (1995).
33. Bort, R., Signore, M., Tremblay, K., Martinez Barbera, J. P. & Zaret, K. S. Hex homeobox gene controls the transition of the endoderm to a pseudostratified, cell emergent epithelium for liver bud development. *Dev. Biol.* **290**, 44–56 (2006).
34. Bhushan, A. *et al.* *Fgf10* is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. *Development* **128**, 5109–5117 (2001).
35. Norgaard, G. A., Jensen, J. N. & Jensen, J. FGF10 signaling maintains the pancreatic progenitor cell state revealing a novel role of Notch in organ development. *Dev. Biol.* **264**, 323–338 (2003).
36. Hart, A., Papadopoulou, S. & Edlund, H. Fgf10 maintains notch activation, stimulates proliferation, and blocks differentiation of pancreatic epithelial cells. *Dev. Dyn.* **228**, 185–193 (2003).
37. Papadopoulou, S. & Edlund, H. Attenuated Wnt signaling perturbs pancreatic growth but not pancreatic function. *Diabetes* **54**, 2844–2851 (2005).
38. Murtaugh, L. C., Law, A. C., Dor, Y. & Melton, D. A. β -catenin is essential for pancreatic acinar but not islet development. *Development* **132**, 4663–4674 (2005).
39. Heiser, P. W., Lau, J., Taketo, M. M., Herrera, P. L. & Hebrok, M. Stabilization of β -catenin impacts pancreas growth. *Development* **133**, 2023–2032 (2006).
40. Temple, S. & Raff, M. C. Clonal analysis of oligodendrocyte development in culture: evidence for a developmental clock that counts cell divisions. *Cell* **44**, 773–779 (1986).
41. Weissman, I. L., Anderson, D. J. & Gage, F. Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. *Annu. Rev. Cell Dev. Biol.* **17**, 387–403 (2001).
42. Dor, Y., Brown, J., Martinez, O. I. & Melton, D. A. Adult pancreatic β -cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429**, 41–46 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. MacDonald, R. DePinho, G. Fishman, H. Bujard and C. Wright for sharing mouse strains, C. Wright for sharing antibodies, and A. Nagy for providing ES cells. We also thank M. Stolovich-Rein and Y. Dor for assistance in quantifying chimaerism, A. Greenwood for assistance with pancreatic lineage quantification, and A. Panikkar for assistance with morphometric analysis. We are grateful to J. Rajagopal, D. Huangfu, Y. Dor and Q. Zhou for critically reading the manuscript, and to S. Snapper, K. Eggan and H. Akutsu for suggestions and technical expertise. L. Peebles and J. Zhang provided assistance with the statistical analysis. B.Z.S. is supported by an award from NIDDK. D.A.M. is an investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.M. (dmelton@harvard.edu) or B.Z.S. (bstanger@mail.med.upenn.edu).

LETTERS

A spectrum of an extrasolar planet

L. Jeremy Richardson¹, Drake Deming², Karen Horning³, Sara Seager^{4,5} & Joseph Harrington⁶

Of the over 200 known extrasolar planets, 14 exhibit transits in front of their parent stars as seen from Earth. Spectroscopic observations of the transiting planets can probe the physical conditions of their atmospheres^{1,2}. One such technique^{3,4} can be used to derive the planetary spectrum by subtracting the stellar spectrum measured during eclipse (planet hidden behind star) from the combined-light spectrum measured outside eclipse (star + planet). Although several attempts have been made from Earth-based observatories, no spectrum has yet been measured for any of the established extrasolar planets. Here we report a measurement of the infrared spectrum (7.5–13.2 μm) of the transiting extrasolar planet HD 209458b. Our observations reveal a hot thermal continuum for the planetary spectrum, with an approximately constant ratio to the stellar flux over this wavelength range. Superposed on this continuum is a broad emission peak centred near 9.65 μm that we attribute to emission by silicate clouds. We also find a narrow, unidentified emission feature at 7.78 μm . Models of these ‘hot Jupiter’⁵ planets predict a flux peak^{6–9} near 10 μm , where thermal emission from the deep atmosphere emerges relatively unimpeded by water absorption, but models dominated by water fit the observed spectrum poorly.

We observed the HD 209458b system during two predicted secondary eclipse events, on 6 and 13 July 2005. For each event, we observed continuously for 6 h, centred on the three-hour duration of the eclipse. We used the Infrared Spectrograph (IRS)¹⁰ on the Spitzer Space Telescope¹¹ in staring mode with the SL1 slit (short wavelength, low resolution), which gives a wavelength coverage of ~ 7.4 – $14.5 \mu\text{m}$ and a spectral resolution ($\lambda/\Delta\lambda$) of 60–120. We analysed a total of 560 individual spectra of the system (280 per eclipse event), each with integration time 60.95 s, in order to obtain a single spectrum of the planet for each event.

Our technique^{3,4} exploits the timing of the eclipse to derive the planetary spectrum from the eclipse depth versus wavelength. Our analysis effectively uses IRS as a multi-channel photometer by searching for the eclipse in the time series of flux at each wavelength. This method is equivalent to subtracting the in-eclipse spectra (planet hidden) from the out-of-eclipse spectra (both star and planet visible). We developed a custom procedure to extract flux spectra from the IRS images, and we verified that our results are robust with respect to the details of this spectral extraction. The Supplementary Information presents a complete discussion of our methodology.

We first verify that the eclipse is visible in the wavelength-integrated flux, as shown in Fig. 1. The eclipse is clearly visible in the total flux, appearing as a dip centred on phase 0.5. The depth of the eclipse is not easily determined from this plot alone, owing to three systematic effects; these are noted in Fig. 1 legend and their corrections are discussed in the Supplementary Information.

To derive the planetary spectrum, we use a differential analysis. Recasting the 280 spectra per eclipse as flux versus time at each

wavelength, we divide by the average spectrum and subtract the average time series to produce residual fluxes. This subtracts two of the systematic effects (the baseline ramp and the telescope pointing oscillation; see Fig. 1 legend and the Supplementary Information). We then fit a model eclipse curve to the time series of residual fluxes at each wavelength; the amplitudes from these fits comprise the planetary spectrum. The remaining systematic effect, a slow drift in telescope pointing, is corrected by our calibration procedure, placing the spectrum in contrast units (ratio of planetary flux to stellar flux).

Figure 2a shows the derived planetary spectra from both eclipse events separately, which allows us to confirm the reality of spectral structure. Two real spectral features are present above the noise level and are seen in both eclipse events: (1) a broad feature centred near 9.65 μm ; and (2) a sharp feature occupying only a few wavelength channels centred near 7.78 μm (which is confirmed by a separate

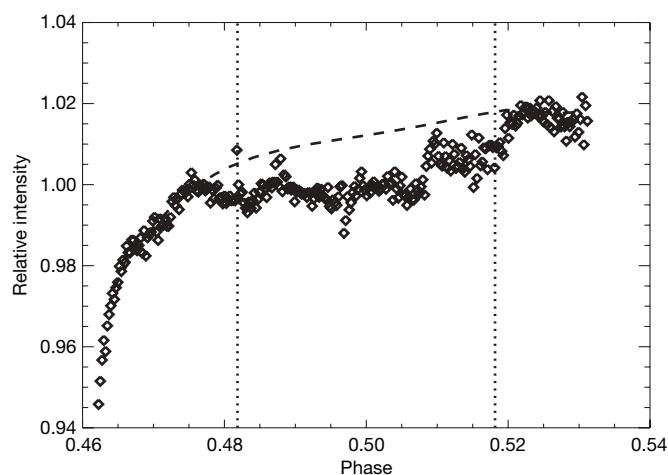


Figure 1 | Wavelength-integrated flux as a function of orbital phase, showing the detection of the secondary eclipse, centred at phase 0.5. The plot shows the total flux, calculated by summing all wavelengths for each spectrum. The results for the two eclipse events are then added together and normalized to the mean value of the total flux of both events. The eclipse (with apparent depth $\sim 0.5\%$) is observable in spite of several systematic effects. The known boundaries of the eclipse (first and fourth contacts) as derived from data in the visible^{28,29} are indicated by the vertical dotted lines. Three systematic effects (see Supplementary Information for details) are removed by our analysis and are present in this figure: (1) a slow ramp-up in intensity of the baseline³⁰ (dashed curve); (2) a telescope pointing oscillation of 1.02-h period that modulates the flux transmitted through the instrument slit (although this effect is difficult to see here, as the oscillation was nearly out of phase for the two eclipses); and (3) a slow drift in telescope pointing that causes an extra dip in intensity and adds to the apparent depth of the eclipse.

¹Exoplanets and Stellar Astrophysics Laboratory, Mail Code 667, ²Planetary Systems Laboratory, Mail Code 693, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ³Department of Physics and Space Sciences, Florida Institute of Technology, 150 West University Boulevard, Melbourne, Florida 32901, USA. ⁴Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Rd, NW, Washington DC 20015, USA. ⁵Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ⁶Department of Physics, University of Central Florida, Orlando, Florida 32816, USA.

analysis shown in Fig. 3 and detailed in the Supplementary Information). Both of these spectral features appear in emission, that is, in excess of the apparent continuum level. Figure 2b shows the average of the two eclipse events. A χ^2 analysis confirms the presence of structure in the spectrum. Specifically, a flat line (that is, constant contrast) is inconsistent with the data at the 3.5σ level. Recall that the eclipse is seen clearly in Fig. 1. After correcting for systematic errors, the eclipse depth ($\sim 0.3\%$) exceeds the errors at individual wavelengths in Fig. 2. Therefore, a flat line in Fig. 2 would also represent a clearly detected, but structureless, spectrum. The reality of the

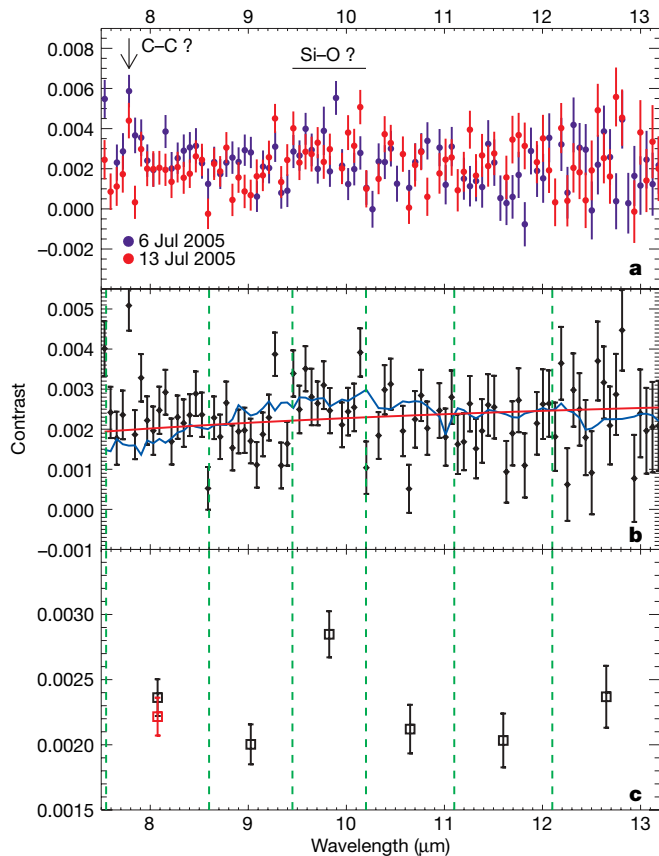


Figure 2 | The spectrum of HD 209458b from 7.5 μm to 13.2 μm . **a**, The result from both observed eclipse events separately. Emission features (near 7.8 μm and 9.65 μm) and candidate identifications are indicated. The units of the y axis are contrast (planet relative to star), and the absolute depth of the eclipse has been calibrated to the preliminary IRAC result at 8 μm (D. Charbonneau, personal communication); see Supplementary Information for details. We make no claims about the spectrum beyond $\sim 10.5 \mu\text{m}$, where the errors increase owing to the decreasing flux and the points are correspondingly more scattered. **b**, The average of the two events with models overplotted. The blue curve is a model for HD 209458b¹² (which is consistent with the photometric result at 24 μm ; ref. 13); the red curve is a 1,600 K blackbody for the planet divided by a 6,000 K blackbody for the star (although a range of blackbody temperatures for the planet from 1,100 K to 1,600 K were tested). **c**, A coarse binning in wavelength (boundaries indicated by green dashed vertical lines) of the average spectrum from **b**. The bins were defined to probe the spectral features discussed in the main text. The weighted mean of all the points in each bin is calculated, and the error on the mean is also weighted by the errors on the individual points. An average of 14 data points appear in each bin. For the bin at the shortest wavelength, two points are shown: one including the 7.8 μm emission feature (black) and one excluding this feature (red). In **a–c**, error bars represent \pm s.e.m.; that is, they are calculated by propagating the errors in the individual points to determine the error on the mean. Also, we show in **a–c** only the result shortward of 13.2 μm because IRS spectra at the longest wavelengths are affected by a systematic error called the ‘teardrop effect’. This effect is not well understood, but is believed to be caused by scattered light (see the IRS Data Handbook v.2.0, p. 62).

broad feature at 9.65 μm is further illustrated by Fig. 2c. This plot shows the average spectrum from Fig. 2b, binned coarsely over wavelength. The rise in flux in the region between 9 and 10 μm is clear and statistically significant (3.6σ difference between flux points at 9 and 10 μm). Several other suggestive features are apparent in Fig. 2a (for example, possible absorption at 8.6 and 9.3 μm), but these are not clearly detected.

We now consider interpretations of the two features observed in the measured spectrum (and summarized in Table 1). First, the 9.65- μm peak (most noticeable as a rise in the spectrum from 9 to 10 μm , as shown in Fig. 2c) is significant at the 3.6σ level when suitably binned to the apparent width of the feature. Because of this peak and the relatively flat spectrum at 10–13 μm , blackbody spectra (in the temperature range 1,100–1,600 K) are ruled out to the $\sim 3.5\sigma$ level. A seemingly natural interpretation of this feature is water vapour absorption at 7–9 μm . Such an absorption feature is prominent in most published HD 209458b models⁶. All hot Jupiter spectra are expected to be shaped by water absorption because water is an abundant gas at the high temperatures of hot Jupiters (1,000–2,000 K). However, we do not favour this water absorption interpretation. We previously reported an upper limit on the water vapour absorption feature at 2.2 μm for the spectrum of this planet⁴. Moreover, on the basis of Fig. 2 alone, a typical solar abundance model of HD 209458b with strong water features¹² is ruled out at the 3.5σ level, owing to the poor fit to the spectral slope at the shortest wavelengths. Our results for the contrast in this spectral region are consistent with the depth of the secondary eclipse at much longer wavelength (24 μm)¹³.

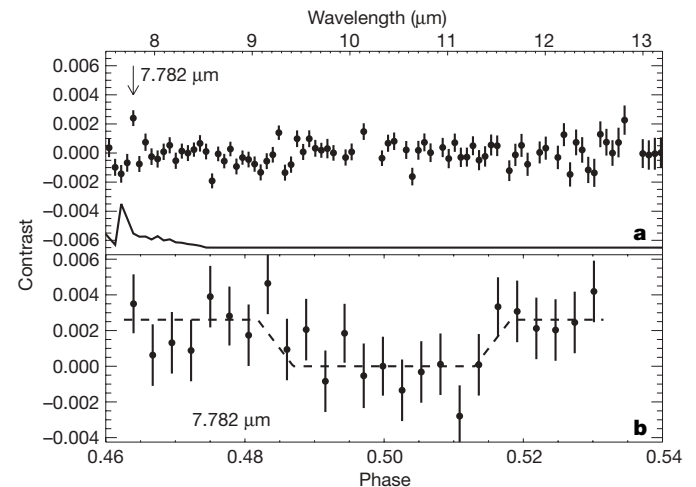


Figure 3 | Separate analysis to confirm the unidentified emission feature near 7.8 μm . As the primary instrument systematics vary slowly with wavelength, an alternative method to eliminate them is to apply a high-pass filter to the observed spectra (which also suppresses any broad planetary spectral structure). We fit a sixth-order polynomial to each spectrum and subtract this fit. Forming a time series of the filtered differences at each wavelength, we fit the model eclipse curve as before. The resulting strongly filtered planetary spectrum, averaged for the two events, is shown in **a**. The solid line is a model of methane line strengths from HITRAN²⁴ scaled to 1,500 K; units are arbitrary and illustrate relative line strengths. The 7.782 μm point stands out, as shown in **a** by an arrow. The point is detected at the 5.4σ significance level, calculated in the same way as described in Table 1 but using the high-pass filtered spectra; the significance level is higher because the wings of the feature have been suppressed. Panel **b** shows the binned time series for the single wavelength channel at 7.782 μm . The model eclipse fit to this time series is overplotted (dashed line), indicating that the differential eclipse is visible at this wavelength with the correct duration and central phase. The error bars in **a** and **b** represent \pm s.e.m., calculated by propagating the errors from the individual points to determine the error on the mean.

Table 1 | Detected features in the spectrum of HD 209458b

Lower limit, λ_{\min} (μm)	Upper limit, λ_{\max} (μm)	Number of channels	Average contrast	s.e.m.	Significance level (σ)	Candidate identification
~ 7.65	~ 7.92	~ 2	0.0027	0.00063	4.4	C–C ?
~ 9.3	~ 10.1	~ 13	0.00085	0.00023	3.6	Si–O

For the narrow feature at 7.8 μm , the width is estimated by fitting the shape of the feature, as described in the Supplementary Information. The average contrast is computed by the taking the single average point at 7.8 μm and subtracting the mean of the 2 pixels on both sides of the peak (4 pixels total). For the broad feature, the width is only a rough estimate, as we do not claim that the feature has a definite downturn beyond 10.5 μm . The average contrast and error are based on the binned spectrum in Fig. 2c. Here we take the 'peak' bin near 10 μm and subtract the 'continuum' bin at 9 μm to get the average contrast, and the error is the relative error between the two points. For both features, the significance level is simply the average contrast divided by the s.e.m.

The occurrence of a peak at 9.65 μm is strongly reminiscent of the Si–O fundamental stretching mode¹⁴ at 9.7 μm , manifested in this case as silicate clouds. Absorption and emission from amorphous and crystalline silicates are ubiquitous in young star- and planet-forming regions¹⁵, and silicates can also condense directly in hot Jupiter atmospheres^{16,17}. Recent observations¹⁸ of L dwarfs reveal 10 μm absorption by silicate clouds. The silicate grains must be small ($<10 \mu\text{m}$) to exhibit the feature¹⁸, suggesting that they can occur at high altitudes. Further, to produce a silicate feature in emission requires that silicate clouds be present in a region of inverted temperature gradient. We hypothesize that the feature could be explained by high silicate clouds in the stratosphere with an inverted temperature gradient. Several recent studies have suggested the possibility of a deep stratosphere on hot Jupiters. The discovery of OGLE-TR-56b prompted models that include strong stellar irradiation, and one study concluded that TiO in the upper atmosphere can cause a temperature inversion¹⁹. More recently, the detection of thermal emission from TrES-1²⁰ using IRAC revealed a higher brightness temperature at 8 μm than at 4.5 μm , which was unexpected on the basis of previous models, and one of several explanations is a temperature inversion^{21,22}. Finally, in this respect, we note that the presence of high clouds (to \sim millibar pressures) is consistent with other observational results for HD 209458b, specifically the low sodium abundance¹, the upper limit on CO absorption during transit², and the non-detection of water bands in the near-infrared⁴. Unanticipated sources of opacity may be required to produce a temperature inversion at these altitudes, and thereby mask the effect of water opacity.

Alternatively, HD 209458b is known to have an extended and evaporating atmosphere²³, and it is possible that an optically thin, emitting dust envelope could contribute to the 9.65 μm feature. We also warn that our silicate feature is based on a rise in the spectrum near the Si–O stretching resonance, and at the level of data uncertainty we do not claim a downturn beyond 10 μm that would support the silicate feature claim.

The second feature in our spectrum is a narrow, sharp peak at 7.78 μm . This peak is statistically significant at the 4.4σ level and is unlikely to be an instrumental error because the peak appears in the spectra from both observed eclipse events. If produced by thermal emission, this feature is also consistent with an inverted temperature gradient. We considered the possibility that this peak is due to methane emission. Figure 3a includes a profile of the wavelength dependence of methane emission, obtained by scaling the HITRAN²⁴ line strengths to $T = 1,500 \text{ K}$, binning them to IRS resolution, and assuming optically thin emission. The observed peak is not coincident with the strongest methane lines (Q branch). The predicted position of the Q branch shifts to longer wavelength with increasing temperature, but a two-pixel discrepancy remains at any plausible temperature, and a wavelength calibration error of this magnitude is out of the question (Spitzer Support, personal communication). Although other methane features occur over the range extending from ~ 7.4 to $\sim 8.0 \mu\text{m}$ (ref. 24), it seems unlikely that these weaker lines alone could cause the feature in the observed spectrum. A more exotic possibility that cannot be firmly rejected is the C–C stretching resonance in polycyclic aromatic hydrocarbons²⁵. Additional Spitzer observations should clarify the nature of this emission.

Finally, we look forward to future extension of extrasolar planet spectroscopy to the domain of transiting terrestrial planets. Although

the modest size of the Spitzer Space Telescope currently limits us to the brightest transiting planet systems, the 6.5-m aperture of the forthcoming James Webb Space Telescope²⁶ should provide a sufficient photon flux to measure the spectrum of a transiting 'hot Earth' orbiting a nearby lower-main-sequence star²⁷.

Received 11 December 2006; accepted 1 February 2007.

- Charbonneau, D., Brown, T. M., Noyes, R. W. & Gilliland, R. L. Detection of an extrasolar planet atmosphere. *Astrophys. J.* **568**, 377–384 (2002).
- Deming, D., Brown, T. M., Charbonneau, D., Harrington, J. & Richardson, L. J. A new search for carbon monoxide absorption in the transmission spectrum of the extrasolar planet HD 209458b. *Astrophys. J.* **622**, 1149–1159 (2005).
- Richardson, L. J. *et al.* Infrared observations during the secondary eclipse of HD 209458b. I. 3.6 Micron occultation spectroscopy using the Very Large Telescope. *Astrophys. J.* **584**, 1053–1062 (2003).
- Richardson, L. J., Deming, D. & Seager, S. Infrared observations during the secondary eclipse of HD 209458b. II. Strong limits on the infrared spectrum near 2.2 microns. *Astrophys. J.* **597**, 581–589 (2003).
- Collier Cameron, A. Extrasolar planets: What are hot Jupiters made of? *Astron. Geophys.* **43**, 21–24 (2002).
- Marley, M. S., Fortney, J., Seager, S. & Barman, T. Atmospheres of extrasolar giant planets. Preprint at (<http://arxiv.org/astro-ph/0602468>) (2006).
- Burrows, A. A theoretical look at the direct detection of giant planets outside the Solar System. *Nature* **433**, 261–268 (2005).
- Sudarsky, D., Burrows, A. & Hubeny, I. Theoretical spectra and atmospheres of extrasolar giant planets. *Astrophys. J.* **588**, 1121–1148 (2003).
- Seager, S. & Sasselov, D. D. Extrasolar giant planets under strong stellar irradiation. *Astrophys. J.* **502**, L157–L161 (1998).
- Houck, J. R. *et al.* The Infrared Spectrograph (IRS) on the Spitzer Space Telescope. *Astrophys. J.* **154** (Suppl.), 18–24 (2004).
- Werner, M. W. *et al.* The Spitzer Space Telescope Mission. *Astrophys. J.* **154** (Suppl.), 1–9 (2004).
- Seager, S. *et al.* On the dayside thermal emission of hot Jupiters. *Astrophys. J.* **632**, 1122–1131 (2005).
- Deming, D., Seager, S., Richardson, L. J. & Harrington, J. Infrared radiation from an extrasolar planet. *Nature* **434**, 740–743 (2005).
- Dorschner, J. Infrared spectra of silicate grains. *Astron. Nachr.* **293**, 53–55 (1971).
- Kessler-Silacci, J. *et al.* C2D Spitzer IRS spectra of disks around T Tauri stars. I. Silicate emission and grain growth. *Astrophys. J.* **639**, 275–291 (2006).
- Burrows, A. & Sharp, C. M. Chemical equilibrium abundances in brown dwarf and extrasolar giant planet atmospheres. *Astrophys. J.* **512**, 843–863 (1999).
- Seager, S., Whitney, B. A. & Sasselov, D. D. Photometric light curves and polarization of close-in extrasolar giant planets. *Astrophys. J.* **540**, 504–520 (2000).
- Cushing, M. C. *et al.* A Spitzer Infrared Spectrograph spectral sequence of M, L, and T dwarfs. *Astrophys. J.* **648**, 614–628 (2006).
- Hubeny, I., Burrows, A. & Sudarsky, D. A possible bifurcation in atmospheres of strongly irradiated stars and planets. *Astrophys. J.* **594**, 1011–1018 (2003).
- Charbonneau, D. *et al.* Detection of thermal emission from an extrasolar planet. *Astrophys. J.* **626**, 523–529 (2005).
- Fortney, J. J., Marley, M. S., Lodders, K., Saumon, D. & Freedman, R. Comparative planetary atmospheres: Models of TrES-1 and HD 209458b. *Astrophys. J.* **627**, L69–L72 (2005).
- Fortney, J. J., Saumon, D., Marley, M. S., Lodders, K. & Freedman, R. S. Atmosphere, interior, and evolution of the metal-rich transiting planet HD 149026b. *Astrophys. J.* **642**, 495–504 (2006).
- Vidal-Madjar, A. *et al.* An extended upper atmosphere around the extrasolar planet HD209458b. *Nature* **422**, 143–146 (2003).
- Rothman, L. S. *et al.* The HITRAN 2004 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **96**, 139–204 (2005).
- Sloan, G. C. *et al.* Mid-infrared spectra of polycyclic aromatic hydrocarbon emission in Herbig Ae/Be stars. *Astrophys. J.* **632**, 956–963 (2005).
- Gardner, J. P. *et al.* The James Webb Space Telescope. *Space Sci. Rev.* **123**, 485–606 (2006).
- Tarter, J. C. *et al.* A re-appraisal of the habitability of planets around M dwarf stars. Preprint at (<http://arxiv.org/astro-ph/0609799>) (2006).
- Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W. & Burrows, A. Hubble Space Telescope time-series photometry of the transiting planet of HD 209458. *Astrophys. J.* **552**, 699–709 (2001).

29. Knutson, H., Charbonneau, D., Noyes, R. W., Brown, T. M. & Gilliland, R. L. Using stellar limb-darkening to refine the properties of HD 209458b. *Astrophys. J.* **655**, 564–575 (2007).
30. Deming, D., Harrington, J., Seager, S. & Richardson, L. J. Strong infrared emission from the extrasolar planet HD 189733b. *Astrophys. J.* **644**, 560–564 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. We acknowledge cooperation with D. Charbonneau, C. Grillmair and H. Knutson. Our understanding of the long-term telescope pointing drift was derived from H. Knutson's study of the effect in their 30 h programme to measure the light curve of HD 189733b. D. Charbonneau provided his measurement of the eclipse depth of HD 209485b,

which was key to casting our results in terms of contrast, rather than differential contrast. We also thank M. Swain and A. Mainzer for discussions regarding the telescope pointing oscillation. We thank the teams that designed, built, operate and support the Spitzer Space Telescope and the IRS. We also thank the NASA Astrobiology Institute, which has centres at both NASA Goddard and the Carnegie Institution of Washington. L.J.R. acknowledges support as a NASA Postdoctoral Fellow at NASA Goddard (formerly the NRC Research Associateship Program). K.H. performed the high-pass filtering analysis as part of her participation in the Summer Undergraduate Internship in Astrobiology, funded by the Goddard Center for Astrobiology. S.S. thanks the Spitzer Theory Program and the Carnegie Institution of Washington for support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.J.R. (lee.richardson@colorado.edu).

LETTERS

Quantum nature of a strongly coupled single quantum dot–cavity system

K. Hennessy^{1,2*}, A. Badolato^{1*}, M. Winger^{1*}, D. Gerace¹, M. Atatüre¹, S. Gulde¹, S. Fält¹, E. L. Hu² & A. Imamoglu¹

Cavity quantum electrodynamics (QED) studies the interaction between a quantum emitter and a single radiation-field mode. When an atom is strongly coupled to a cavity mode^{1,2}, it is possible to realize important quantum information processing tasks, such as controlled coherent coupling and entanglement of distinguishable quantum systems. Realizing these tasks in the solid state is clearly desirable, and coupling semiconductor self-assembled quantum dots to monolithic optical cavities is a promising route to this end. However, validating the efficacy of quantum dots in quantum information applications requires confirmation of the quantum nature of the quantum-dot–cavity system in the strong-coupling regime. Here we find such confirmation by observing quantum correlations in photoluminescence from a photonic crystal nanocavity^{3–5} interacting with one, and only one, quantum dot located precisely at the cavity electric field maximum. When off-resonance, photon emission from the cavity mode and quantum-dot excitons is anticorrelated at the level of single quanta, proving that the mode is driven solely by the quantum dot despite an energy mismatch between cavity and excitons. When tuned to resonance, the exciton and cavity enter the strong-coupling regime of cavity QED and the quantum-dot exciton lifetime reduces by a factor of 145. The generated photon stream becomes antibunched, proving that the strongly coupled exciton/photon system is in the quantum regime. Our observations unequivocally show that quantum information tasks are achievable in solid-state cavity QED.

Pursuit of solid-state cavity QED is motivated by the possibility of fixing the emitter location with respect to the cavity mode electric field maximum and of enhancing the emitter–cavity coupling by fabricating cavities with ultrasmall volumes. Substantial progress has been made to this end, culminating in the demonstration of strong coupling of a Cooper pair box to a superconducting transmission line microwave resonator⁶ and of a semiconductor self-assembled quantum dot (QD) to a nanoscale optical cavity mode^{7–9}. However, further progress in QD cavity QED has been partially hindered by the conventional practice of incorporating many QDs at random locations in the cavity, leading to indirect/off-resonant coupling to other QDs overlapping the mode. Recent experiments showing intense cavity emission, and even lasing, when the mode is non-resonant with the QDs^{4,7–10} suggest that observation of quantum effects would require a nanocavity containing a single QD.

Here we demonstrate a new technique to locate the absolute position of buried QDs by atomic force microscopy (AFM) metrology. As the top half of the photonic crystal slab is grown on the QD layer (see Methods), the presence of the QDs gives rise to small but distinctive 1–2 nm hills on the surface that can be detected by AFM. In ways similar to those used to position nanocavities relative to stacks of QDs¹⁰, we were able to position with 30 nm accuracy a nanocavity

relative to one, and only one, QD, aligning it to the electric-field maximum of the cavity mode as shown in Fig. 1a, b. This positioning technique allows us to study the coupled system with unprecedented clarity. Furthermore, in comparison to devices containing multiple QDs, we achieve significantly higher quality factors (Qs), ranging from 12,000 to 30,000 in ~20 deterministically coupled devices. In typical devices, QDs are located at ~90% of the electric-field maximum.

Our approach allows for pre-selection of QDs with desirable spectral properties. In the remainder of this Letter, we follow one such QD that was selected on the basis of a small QD–mode spectral detuning despite having less-than-average QD–mode alignment. The measured photoluminescence spectrum (see Methods) of this QD (Fig. 1c) consisted of a few narrow, isolated excitonic transitions that were promising for nanocavity coupling. We then fabricated^{10,11} the

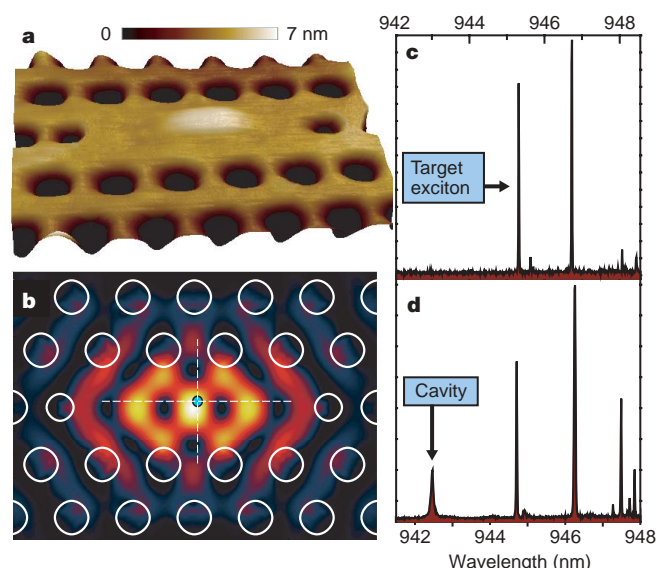


Figure 1 | Positioning a photonic crystal cavity mode relative to a single buried QD. **a**, AFM topography of a photonic crystal nanocavity aligned to a hill of material on the surface arising from a QD buried 63 nm below. The height scale is depicted by the colour bar. **b**, Electric field intensity of the photonic crystal cavity mode showing that the location of the buried QD, indicated by the teal dot, overlaps the field maximum. The field intensity ranges from zero (black) to a maximum (white), going through blue, red and yellow. **c**, Photoluminescence spectrum before cavity fabrication of a single QD, which was selected for cavity coupling on the basis of clear emission from a few discrete excitonic transitions. **d**, Photoluminescence spectrum from the same QD after cavity fabrication, showing emission from the cavity at 942.5 nm.

¹Institute of Quantum Electronics, ETH Zürich, HPT G10, 8093 Zurich, Switzerland. ²California NanoSystems Institute, University of California, Santa Barbara, California 93106, USA. *These authors contributed equally to this work.

photonic crystal with the lattice parameter a and hole radius r/a selected to lithographically tune the nanocavity mode to this precise spectral location. As shown in Fig. 1d, the cavity mode is spectrally positioned a few nanometres shorter in wavelength than the exciton transitions, allowing us to tune the cavity by a thin-film condensation technique¹² and study the exciton–cavity interaction as a function of the detuning between mode and exciton, $\Delta\lambda = \lambda_x - \lambda_m$ ($\Delta\omega = \omega_m - \omega_x$), where λ_x (ω_x) and λ_m (ω_m) denote the resonant exciton and cavity-mode wavelength (angular frequency), respectively.

An important feature of the spectrum in Fig. 1d is that emission from the cavity mode is observed where there was clearly no QD emission before cavity fabrication. Cavity emission for arbitrary detunings challenges the notion that QDs may be fully described as artificial atoms with discrete energy levels. To explain this effect, recent reports speculate that the mode is fed by either a background continuum or a multi-QD effective continuum⁷. We rule out the former by noting that no mode emission occurs for control cavities containing an InAs wetting layer but positively no QDs, and we dismiss the latter by observing that even a single QD sustains efficient mode emission for all detunings observed in our devices ($-19 \text{ nm} < \Delta\lambda < 4.1 \text{ nm}$). We further investigate the nature of the off-resonant coupling by measuring the quantum correlations between photons emitted from the exciton transitions and the spectrally detuned ($\Delta\lambda = 4.1 \text{ nm}$) cavity. This measurement corresponds to the second-order, normally ordered, cross-correlation function¹³ $g_{x,m}^{(2)}(\tau) = \langle I_x(t+\tau)I_m(t) \rangle / \langle I_x(t) \rangle \langle I_m(t) \rangle$, where $I_{x(m)}(t)$ and $I_{x(m)}(t+\tau)$ refer to the intensity of the exciton (cavity-mode) photon stream at time t and a delayed time $t+\tau$, respectively. Arrival of a cavity (exciton) photon triggers a timer that stops upon detection of an exciton (cavity) photon for a positive (negative) difference in arrival time (see Fig. 2a and Methods). We accumulate a histogram of the detection events (Fig. 2b). For zero time difference, we observe strong antibunching, a deviation from the uncorrelated case of $g^{(2)}(0) = 1$ that directly proves the two emission events stem from the same single quantum emitter and are anticorrelated at the level of single quanta. Remarkably, the time constant for negative time delay

(8.5 ns) corresponds to the lifetime of the exciton, $\tau_x = 8.7 \text{ ns}$ (Fig. 2c), while the time constant for positive time delay compares to the lifetime of the mode emission, $\tau_m = 1.3 \text{ ns}$ (Fig. 2d). (This is surprising because we would expect the lifetime of the states producing the stop pulse to dictate the respective time constant.) Off-resonant cavity–exciton anticorrelation demonstrates the existence of a new, unidentified mechanism for channelling QD excitations into a non-resonant cavity mode, indicating a clear deviation from a simple artificial atom model of the QD.

Having studied the exciton–mode coupling for large detuning, we then tuned the mode into resonance with the exciton to study resonant interaction. Clear evidence of the strong-coupling regime of cavity QED is obtained from spectra recorded as a function of $\Delta\lambda$. A subset of these spectra is plotted in Fig. 3b, from which we note the following features: for $\Delta\lambda > 0.05 \text{ nm}$, the short-wavelength spectral feature remains cavity-like with a linewidth of $\Delta\lambda_- \approx \Delta\lambda_m = 0.071 \text{ nm}$ and the long-wavelength peak retains its excitonic nature with $\Delta\lambda_+ \approx \Delta\lambda_x = 0.025 \text{ nm}$. Here, $\Delta\lambda_m$ ($\Delta\lambda_x$) denotes the bare cavity-mode (exciton) linewidth. For further decrease in $\Delta\lambda$, the short-wavelength feature broadens and then separates into two distinct peaks. As tuning proceeds, the middle peak preserves exactly the wavelength, linewidth and polarization of the cavity mode while the long- and short-wavelength peaks repel each other and assume equal linewidths of $\Delta\lambda_{\pm} = 0.042 \text{ nm} \approx (\Delta\lambda_m + \Delta\lambda_x)/2$. Our observation of a spectral triplet in the strong-coupling regime is unique among recent reports of solid-state vacuum Rabi splitting^{7–9}. The two outer peaks at $\Delta\lambda = 0$ anticross, as shown in Fig. 3a, and are identified as the polariton states of the strongly coupled exciton–photon system. Their resonant frequencies Ω_{\pm} and linewidths Γ_{\pm} are:

$$\Omega_{\pm} + i\Gamma_{\pm}/2 = \frac{\omega_m + \omega_x}{2} - i\frac{\gamma_x + \gamma_m}{4} \pm \sqrt{g^2 + \frac{1}{4}\left(\Delta\omega - i\frac{\gamma_x - \gamma_m}{2}\right)^2} \quad (1)$$

where g is the exciton–mode coupling frequency estimated from the minimum observed polariton splitting and the full-width at half-maximum (FWHM) of the mode $\gamma_m = 24.1 \text{ GHz} = 100 \mu\text{eV}$ ($Q \approx 13,300$) and exciton $\gamma_x = 8.5 \text{ GHz} = 35 \mu\text{eV}$. The calculated peak positions are plotted as the continuous lines in Fig. 3a, and show very good agreement with the measured ones. From this Rabi splitting trend, we find $g = 18.4 \text{ GHz} = 76 \mu\text{eV}$ and note that g is reduced to $\sim 70\%$ of its maximum possible value owing to the slight spatial

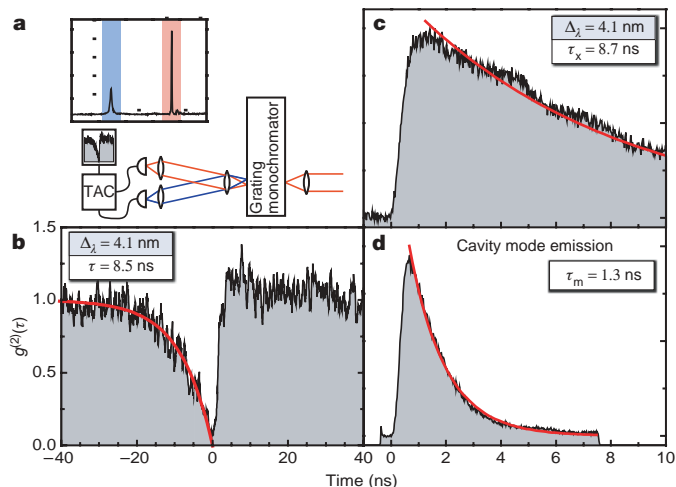


Figure 2 | Cross-correlation histogram and time-resolved photoluminescence from the QD–cavity system with $\Delta\lambda = 4.1 \text{ nm}$. **a**, Diagram of the optical system used in cross-correlation measurements. The mode and exciton wavelengths are filtered as indicated by the blue and red shading, separated by a grating monochromator and imaged on separate single-photon detectors. **b**, Cross-correlation histogram of the mode and exciton emission (black) with exponential fits (red). The two emission events are anticorrelated at the single-photon level, giving rise to an antibunching in $g^{(2)}(\tau)$. **c**, The exciton lifetime $\tau_x = 8.7 \text{ ns}$ is comparable to the time constant observed in the cross-correlation histogram for negative time differences. **d**, The lifetime of the mode emission $\tau_m = 1.3 \text{ ns}$ is comparable to the time constant observed in the cross-correlation histogram for positive time differences.

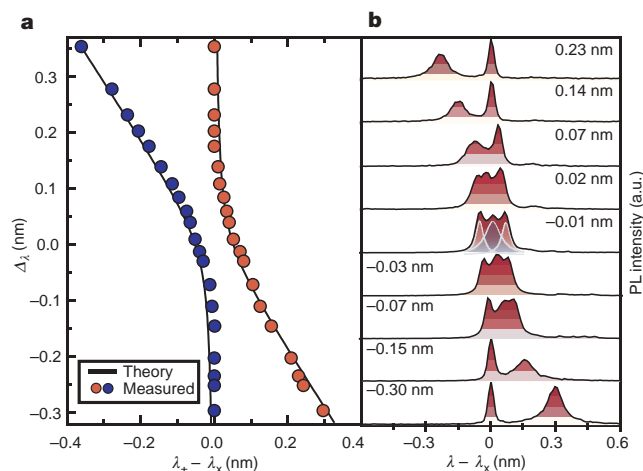


Figure 3 | Characteristics of the strong-coupling regime in the spectral domain. **a**, Wavelength of the polaritons for various detunings, $\Delta\lambda$. Calculated spectral peak positions describing the strongly coupled system are plotted as solid lines, with measured peak positions extracted from photoluminescence plotted in red and blue dots. **b**, Spectra of the two anticrossing polariton states near zero detuning. An additional peak is identified as the pure photonic state of the cavity. Values of $\Delta\lambda$ are shown for each spectrum. PL, photoluminescence; a.u., arbitrary units.

mismatch between the QD and the electric field maximum of the cavity mode. Coupling constants for other instances of strong coupling with different excitons are as high as 120 μeV in our samples.

The predicted strong-coupling spectrum accounts for two of the three peaks that we observe experimentally—we attribute the additional peak to the pure photonic state of the cavity. A similar three-peak lineshape has been observed for Ba atoms strongly coupled to a high-finesse cavity¹⁴, and is attributed to fluctuations in the number of atoms present in the cavity over time. Our data can be regarded as the solid-state analogue of this scenario, in which fluctuations in emitter energy, rather than emitter number, occur over time. As the occupation of charging centres in the vicinity of the QD fluctuates, the exciton energy is renormalized via the Coulomb interaction and the detuning becomes large, $|\Delta_\omega| \gg g$. For such large detunings, the short-wavelength polariton reverts to the pure photonic state of the cavity. It is plausible that the QD exciton emission in this case leads to one of the longer-wavelength photoluminescence lines depicted in Fig. 1d. In our experiments, we integrate the photoluminescence for ~ 1 s, thereby collecting a time-average of the two regimes that results in the observed spectral triplet.

The clear signatures of strong coupling in Fig. 3 are confirmed by measuring τ_x as the mode is tuned into resonance with the exciton. Modification of τ_x is predicted to be particularly strong in a photonic crystal environment owing to the depleted density of optical states for energies in the photonic bandgap¹⁵. This is confirmed by our measurement at large detuning ($\Delta_\lambda = 4.1$ nm) of $\tau_x = 8.7$ ns, which is long in comparison to the lifetime in bulk material, $\tau_0 \approx 1$ ns. On the other hand, as the cavity mode spectrally approaches the exciton, τ_x should decrease owing to an effect that can be considered the counterpart of the Purcell effect^{16,17} for a detuned, strongly coupled system. As the cavity is tuned into resonance, τ_x decreases rapidly to $\tau_x \approx 1.6$ ns at $\Delta_\lambda = 1.26$ nm and further diminishes to a carrier-capture limited $\tau_x = 60$ ps at $\Delta_\lambda = 0$ (Fig. 4a, b). The reduction in lifetime by a factor of 145 confirms that our QD is coupled to the cavity mode and spatially located near the cavity centre. Such short

lifetimes, approaching the cavity photon storage time, support the spectral evidence that the exciton–photon coupling g is sufficiently large for the system to be in the strong-coupling regime. In order to estimate g independently from the Rabi splitting, we assume that τ_x is primarily determined by a lorentzian dependence on the detuning¹⁸, and fit the experimental data (see Methods) to obtain $g = 21.8$ GHz = 90 μeV , which agrees very well with the value obtained from the Rabi splitting.

The spectral anticrossing that we observe is unequivocally a result of strong coupling; however, we emphasize that the calculated two-peak lineshape also describes exactly two coupled classical harmonic oscillators¹⁹. Therefore, Rabi splitting alone is not sufficient to discriminate between a regime arising from the interaction of single exciton and photon quanta and a classical regime describing the coupling behaviour of two classical oscillators. For cavity photon occupation numbers above one, the quantum nonlinear regime is manifested in the Jaynes–Cummings ladder spectrum²⁰, which is absent in the case of two strongly coupled classical oscillators. A direct experimental demonstration of the quantum nonlinear regime has been reported for single atoms flying through microwave cavities²¹ and, in the case of optical cavities, has been inferred by monitoring the statistics of the emitted photon stream²². In QD cavity QED, in which the artificial atom is composed of many atoms embedded in a host matrix, it is crucial to confirm that the system manifests true quantum behaviour. To this end, we measure the second-order auto-correlation function $g_{m,m}^{(2)}(\tau)$ of the cavity photon stream under pulsed excitation. For $\Delta_\lambda = 4.1$ nm, the cavity displays nearly poissonian photon statistics as shown in Fig. 4c, with $g_{m,m}^{(2)}(0) \approx 1$, revealing its harmonicity. That is, the cavity is accepting multiple photons at the same time—a surprising result given the observed $g^{(2)}(0) \approx 0$ in cross-correlation with the exciton. We then tuned the system into strong coupling and measured the auto-correlation function of the entire three-peak structure consisting of the strong-coupling doublet and the pure photonic state of the cavity. In this regime, the statistics of the cavity photon stream dramatically change to sub-poissonian (Fig. 4d), with a central peak area at $\tau = 0$ that is only 54% of the average peak area at other times.

We note that the area of the peak at $\tau = 0$ is increased from the ideal value of zero by two mechanisms. First, the central peak in the observed spectral triplet, which is not antibunched, contributes 45% of the total collected photoluminescence, as determined by the three-lorentzian fit shown in Fig. 3b. Second, in strong coupling, the carrier capture process proceeds on timescales longer than the polariton decay time (13.3 ps), so that it is possible for the system to undergo multiple capture/emission events per excitation pulse. Accounting for these factors, the measured $g^{(2)}(\tau)$ shows clearly that the cavity field attains an anharmonic character when brought into strong coupling with the exciton, proving the interaction is with a two-level quantum system. We have confirmed that the main contribution to the peak at $\tau = 0$ is from the pure photonic state of the cavity by performing auto-correlation measurements on another device that exhibited larger Rabi splitting. In this device, we spectrally filtered a single polariton peak so that the pure photonic state of the cavity contributed only 15% of the total signal (Supplementary Fig. S1). In this case, the auto-correlation signal exhibited strong antibunching with a peak area at $\tau = 0$ of only 19% (Supplementary Fig. S2). These photon correlation measurements prove the quantum nature of the strongly coupled QD–cavity molecule.

A direct demonstration of the photon blockade effect²³ could be obtained by exciting the photonic channel of the strongly coupled system in reflection or transmission measurements, thereby reaching the ultimate limit of solid-state nonlinear optics at the single-photon level. The results of the present work encourage the experimental pursuit of quantum information tasks in the solid state, such as the deterministic coupling of two QDs mediated by a common cavity mode²⁴.

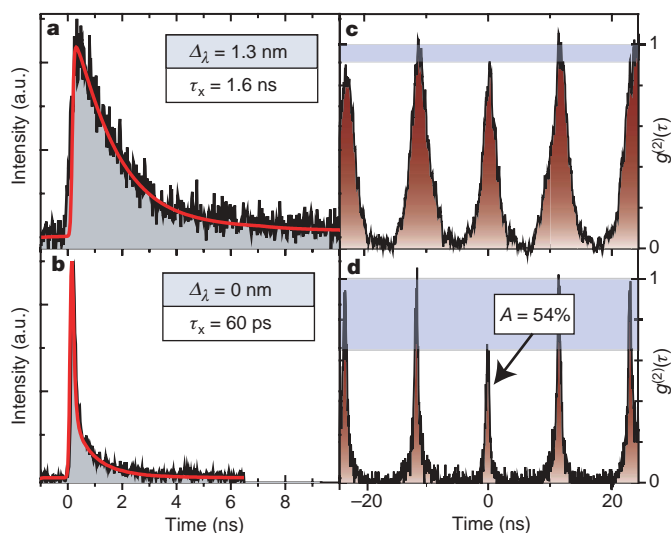


Figure 4 | Characteristics of the strong-coupling regime in the time domain. **a**, The exciton lifetime τ_x is reduced to 1.6 ns when the detuning is $\Delta_\lambda = 1.3$ nm. **b**, At exact spectral resonance, τ_x decreases to 60 ps, a reduction by a factor of 145 from the value at $\Delta_\lambda = 4.1$ nm. **c**, Autocorrelation function of the cavity mode off-resonance with the exciton, showing no significant quantum correlations. **d**, Autocorrelation of the strongly coupled cavity–QD system demonstrating strong quantum correlations in the form of photon antibunching, where the central peak area, A , is 54% of the area of peaks at other times. In **a** and **b**, histogram data are shown in black and an exponential fit in red. In **c** and **d**, the grey bar shows the difference between the central peak height and the average height at other times.

METHODS

Semiconductor material. The semiconductor heterostructure used for subsequent photonic crystal fabrication was grown on a (100) semi-insulating GaAs substrate by molecular beam epitaxy. The epitaxial structure consists of a 126 nm GaAs slab incorporating a layer of InAs quantum dots at the centre with a density ranging from zero to 3 per μm^2 across the sample. By annealing the quantum dots *in situ* when partially capped with GaAs, we tuned the quantum-dot ground state energies to ~ 1.3 eV. The photonic crystal slab is grown above a 1 μm $\text{Al}_{0.7}\text{Ga}_{0.3}\text{As}$ layer that is later removed to provide vertical optical isolation.

Optical characterization. For micro-photoluminescence experiments, samples are mounted in a liquid-helium flow cryostat at 4.2 K. We excite with an above-bandgap ($\lambda = 780$ nm) continuous wave diode laser focused to a 1 μm spot using a $50\times$ microscope objective (NA = 0.55). The photoluminescence signal is collected by the same objective, dispersed by a 750 mm grating spectrograph with a spectral resolution of 21 pm, and detected with a liquid-nitrogen-cooled charge coupled device camera.

In time-dependent lifetime measurements, we excite the sample with a Q-switched pulsed diode laser at a rate of either 40 MHz or 80 MHz. We spectrally filter the photoluminescence using the spectrometer (bandwidth = $0.4\text{ nm} = 550\text{ }\mu\text{eV}$) and focus it on an avalanche photodiode (APD) single-photon counter. A time-to-amplitude converter (TAC) transforms time differences between photon detection events and synchronization pulses provided by the laser to electrical signals of corresponding amplitude. A multichannel analyser accumulates a histogram of the measured intervals. We can choose between two different types of APDs to obtain time resolution of either ~ 70 ps or ~ 400 ps, with the latter having significantly higher quantum efficiency.

In auto-correlation measurements, filtered light is sent to a Hanbury Brown–Twiss interferometer, consisting of a 50:50 plate beam splitter with an APD at either output. Time-correlation of the signals is done in a manner similar to that in lifetime measurements. Cross-correlation measurements are done in the same way after the two wavelengths of interest are spatially separated by the spectrometer.

Estimation of g . We assume that τ_x is primarily determined by a lorentzian dependence on the detuning¹⁸ and fit the experimental data with the relation $\hbar/\tau_x = \gamma_b + \gamma_{\text{SE}}$, in which γ_b is the background emission rate into all other modes and $\gamma_{\text{SE}} = \gamma_m g^2 / [d_\omega^2 + (\gamma_m/2)^2]$ is the spontaneous emission rate into the cavity mode. The best fit is given by $\gamma_b = 12.2\text{ MHz} = 0.05\text{ }\mu\text{eV}$ and $g = 21.8\text{ GHz} = 90\text{ }\mu\text{eV}$. The first value corresponds to a lifetime of 13 ns, strongly modified from the bulk lifetime by the photonic bandgap. The second value confirms that in our system the strong-coupling condition $g^2 > (\gamma_x - \gamma_m)^2/16$ is satisfied and agrees very well with the value of g estimated from the vacuum-field Rabi splitting.

Received 4 October 2006; accepted 8 January 2007.

Published online 28 January 2007.

1. Mabuchi, H. & Doherty, A. C. Cavity quantum electrodynamics: Coherence in context. *Science* **298**, 1372–1377 (2002).
2. Raimond, J. M., Brune, M. & Haroche, S. Colloquium: Manipulating quantum entanglement with atoms and photons in a cavity. *Rev. Mod. Phys.* **73**, 565–582 (2001).

3. Painter, O. *et al.* Two-dimensional photonic band-gap defect mode laser. *Science* **284**, 1819–1821 (1999).
4. Strauf, S. *et al.* Self-tuned quantum dot gain in photonic crystal lasers. *Phys. Rev. Lett.* **96**, 127404 (2006).
5. Akahane, Y., Asano, T., Song, B. S. & Noda, S. High-Q photonic nanocavity in a two-dimensional photonic crystal. *Nature* **425**, 944–947 (2003).
6. Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
7. Peter, E. *et al.* Exciton-photon strong-coupling regime for a single quantum dot embedded in a microcavity. *Phys. Rev. Lett.* **95**, 067401 (2005).
8. Yoshie, T. *et al.* Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity. *Nature* **432**, 200–203 (2004).
9. Reithmaier, J. P. *et al.* Strong coupling in a single quantum dot-semiconductor microcavity system. *Nature* **432**, 197–200 (2004).
10. Badolato, A. *et al.* Deterministic coupling of single quantum dots to single nanocavity modes. *Science* **308**, 1158–1161 (2005).
11. Hennessy, K., Badolato, A., Petroff, P. M. & Hu, E. L. Positioning photonic crystal cavities to single InAs quantum dots. *Photonics Nanostruct. Fund. Applic.* **2**, 65–72 (2004).
12. Strauf, S. *et al.* Frequency control of photonic crystal membrane resonators by monolayer deposition. *Appl. Phys. Lett.* **88**, 043116 (2006).
13. Kiraz, A. *et al.* Photon correlation spectroscopy of a single quantum dot. *Phys. Rev. B* **65**, 161303 (2002).
14. Childs, J. J., An, K., Otteson, M. S., Dasari, R. R. & Feld, M. S. Normal mode line shapes for atoms in standing-wave optical resonators. *Phys. Rev. Lett.* **77**, 2901–2904 (1996).
15. Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
16. Purcell, E. M. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69**, 681 (1946).
17. Englund, D. *et al.* Controlling the spontaneous emission rate of single quantum dots in a two-dimensional photonic crystal. *Phys. Rev. Lett.* **95**, 013904 (2005).
18. Andreani, L. C., Panzarini, G. & Gerard, J. M. Strong-coupling regime for quantum boxes in pillar microcavities: Theory. *Phys. Rev. B* **60**, 13276–13279 (1999).
19. Pau, S., Bjork, G., Jacobson, J., Cao, H. & Yamamoto, Y. Microcavity exciton-polariton splitting in the linear regime. *Phys. Rev. B* **51**, 14437–14447 (1995).
20. Haroche, S. *Fundamental Systems in Quantum Optics* (Elsevier, New York, 1992).
21. Brune, M. *et al.* Quantum Rabi oscillation: A direct test of field quantization in a cavity. *Phys. Rev. Lett.* **76**, 1800–1803 (1996).
22. Birnbaum, K. M. *et al.* Photon blockade in an optical cavity with one trapped atom. *Nature* **436**, 87–90 (2005).
23. Imamoglu, A., Schmidt, H., Woods, G. & Deutsch, M. Strongly interacting photons in a nonlinear cavity. *Phys. Rev. Lett.* **79**, 1467–1470 (1997).
24. Imamoglu, A. *et al.* Quantum information processing using quantum dot spins and cavity QED. *Phys. Rev. Lett.* **83**, 4204–4207 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support by the Swiss National Research Foundation through the ‘Quantum Photonics NCCR’.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.I. (imamoglu@phys.ethz.ch).

LETTERS

Enantioselective halocyclization of polyprenoids induced by nucleophilic phosphoramidites

Akira Sakakura¹, Atsushi Ukai¹ & Kazuaki Ishihara¹

Polycyclic bio-active natural products that contain halogen atoms have been isolated from a number of different marine organisms¹. The biosynthesis of these natural products appears to be initiated by an electrophilic halogenation reaction at a carbon–carbon double bond^{2–4} via a mechanism that is similar to a proton-induced olefin polycyclization^{5–8}. Enzymes such as haloperoxidases generate an electrophilic halonium ion (or its equivalent), which reacts with the terminal carbon–carbon double bond of the polyprenoid, enantioselectively inducing a cyclization reaction that produces a halogenated polycyclic terpenoid. Use of an enantioselective halocyclization reaction is one possible way to chemically synthesize these halogenated cyclic terpenoids; although several brominated cyclic terpenoids have been synthesized via a diastereoselective halocyclization reaction that uses stoichiometric quantities of a brominating reagent^{9–12}, the enantioselective halocyclization of isoprenoids induced by a chiral promoter has not yet been reported. Here we report the enantioselective halocyclization of simple polyprenoids using a nucleophilic promoter. Achiral nucleophilic phosphorus compounds are able to promote the diastereoselective halocyclization reaction to give a halogenated cyclic product in excellent yields. Moreover, chiral phosphoramidites promote the enantioselective halocyclization of simple polyprenoids with *N*-iodosuccinimide to give iodinated cyclic products in up to 99% enantiomeric excess and diastereomeric excess. To the best of our knowledge, this is the first successful example of the enantioselective halopolycyclization of polyprenoids.

To achieve high enantioselectivity in electrophilic halocyclization, we focused on the method used to activate the halogenating reagents by chiral promoters (Fig. 1). In the Lewis acid approach (method A)^{13,14}, the activated halogen atom (*X*) would be placed far from the chiral environment of the Lewis acid (*LA*⁺) in the active species, and therefore it is difficult to obtain high enantioselectivity. In contrast, if chiral nucleophilic promoters (*Nu*⁺) can activate the halogenating reagents (method B), *X* could be placed close to the chiral environment of the nucleophilic promoter in the active species, and the halogenated products would be obtained with high enantioselectivity. However, no successful methods for enantioselective halocyclizations of polyprenoids induced by nucleophilic promoters have been reported, although there have been several examples of halocyclizations using chiral amine promoters with quite low enantioselectivity^{15–18}. Therefore, we initially investigated nucleophilic promoters for the activation of halogenating reagents.

First, we examined the activities of various achiral nucleophiles that promoted the halocyclization of 4-(homogeranyl)toluene (**1**) with *N*-iodosuccinimide (NIS) or *N*-bromosuccinimide (NBS). The reaction was catalytically conducted with one equivalent of NIS or NBS in the presence of 30 mol.% of nucleophilic promoter in dichloromethane (CH₂Cl₂) (Table 1). The reaction gave the desired *trans*-fused AB-ring products **2**, together with *endo*- and *exo*-isomeric

A-ring products **3** (2:3 = ~7:3 in all cases). *cis*-A-ring products and *cis*-fused AB-ring products were not obtained. Fortunately, treatment of a mixture of **2** and **3** with chlorosulphonic acid (ClSO₃H) in 2-nitropropane (*i*-PrNO₂) gave **2** as a single *trans*-diastereomer⁵. Therefore, the yield of **2** was evaluated after subsequent diastereoselective cyclization with ClSO₃H. As a result, tributylphosphine (PBu₃), triphenylphosphine (PPh₃), tris(*p*-methoxyphenyl)phosphine (P[C₆H₄(*p*-OMe)]₃), and triphenyl phosphite [P(OPh)₃] showed good catalytic activities (entries 1–4 in Table 1). In particular, the use of highly nucleophilic tributylphosphine gave the best results (99% for NIS and 81% for NBS, entry 1). When the reaction was conducted in toluene, the catalytic activity of triphenylphosphine unexpectedly decreased (entry 3). NIS and NBS showed comparable reactivity in the present reaction. In contrast, unexpectedly, nucleophilic amines such as 1,4-diazabicyclo[2.2.2]octane (DABCO) and 4-(*N,N*-dimethylamino)pyridine (DMAP) were completely inert (entries 5 and 6). Thus, nucleophilic phosphorus compounds catalysed the electrophilic iodination/bromination of the terminal olefin of **1**, and this efficiently induced diastereoselective cyclization to give the halogenated tricyclic product **2** as a single diastereomer.

Encouraged by these results, we next investigated the enantioselective iodocyclization of **1** with NIS using chiral nucleophilic promoters (Table 2). Commercially available chiral phosphoramidite **13**

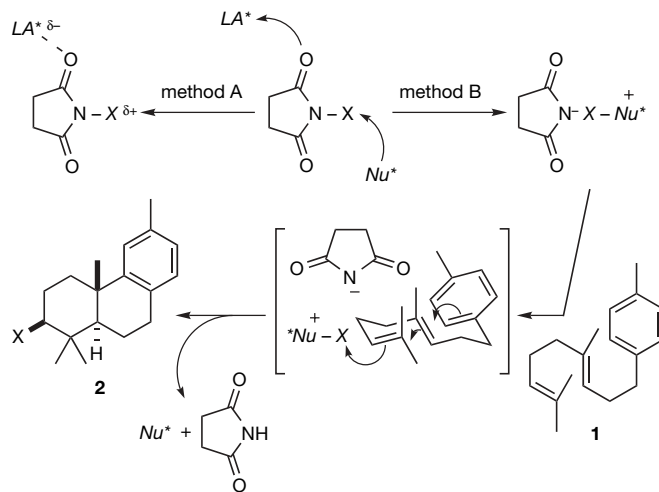
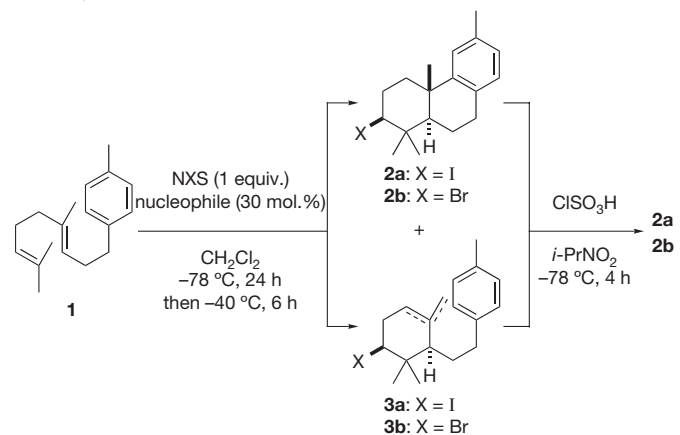


Figure 1 | Nucleophilic promoters versus Lewis acid promoters for the activation of *N*-halosuccinimides, and halocyclization of 4-(homogeranyl)toluene (1**) to **2**.** *LA*⁺ represents a chiral Lewis acid promoter; *Nu*⁺ represents a chiral nucleophilic promoter; *X* represents an activated halogen atom. In the nucleophilic promoter approach, *X* is expected to be placed close to the chiral environment of *Nu*⁺ in the active species, while *X* would be placed far from the chiral environment of *LA*⁺ in the Lewis acid approach.

¹Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa, Nagoya 464-8603, Japan.

Table 1 | Activities of achiral nucleophiles for halopolycyclization of 1.

Entry	Nucleophile	Yield of 2 (%)	
		X = I	X = Br
1	PBu_3	99	81
2	$\text{P}[\text{C}_6\text{H}_4(p\text{-OMe})]_3$	87	88
3	PPh_3	67 (0)*	71 (0)*
4	$\text{P}(\text{OPh})_3$	51	60
5	DABCO	0	0
6	DMAP	0	0
7	No catalyst	3	0

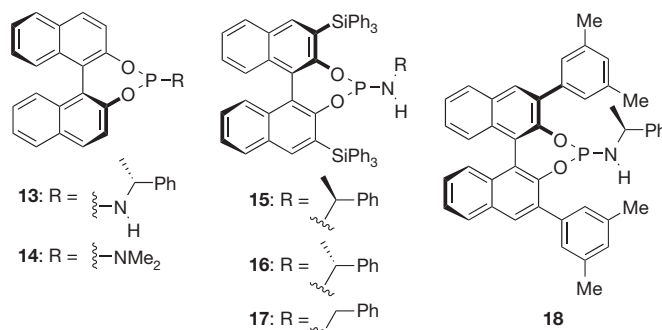
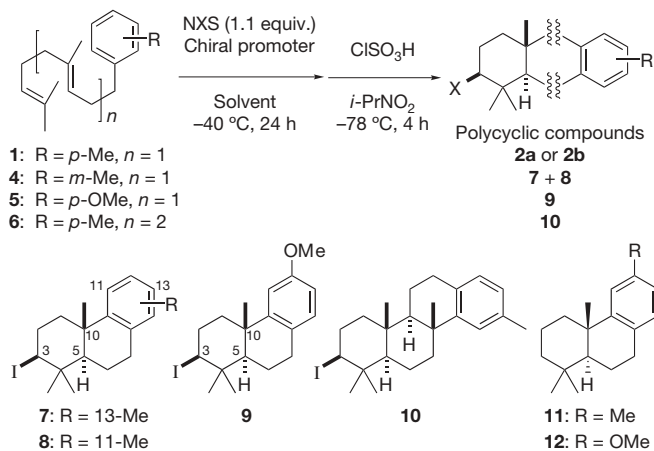
Unless otherwise noted, the reaction of **1** (0.1 mmol) and *N*-halosuccinimide (NXS, 1 equivalent) in CH_2Cl_2 (1 ml) was conducted in the presence of a nucleophile (30 mol.%) at -78°C for 24 h and then at -40°C for 6 h. The conversion yield of **2** was evaluated by ^1H nuclear magnetic resonance analysis after a mixture of **2** and **3** was treated with ClSO_3H (50 μl) in 2-nitropropane (2 ml) at -78°C for 4 h.

* The reaction was conducted in toluene.

(30 mol.%), which is a phosphoramidite of a primary amine, showed high activity (95% yield) in CH_2Cl_2 (entry 1). Unfortunately, the obtained **2a** was racemic. When the reaction was conducted in toluene, **2a** was obtained in 11% yield with 0% enantiomeric excess, even in the presence of 100 mol.% of **13** (entry 2). Commercially available chiral phosphoramidite **14**, which is a phosphoramidite of a secondary amine, showed lower activity (46% yield) in CH_2Cl_2 than **13** (entry 3). The phosphoramidite of a primary amine seemed to be important for good activity.

Next, we designed the chiral phosphoramidite **15**, which has two triphenylsilyl groups at the 3 and 3' positions. Despite the bulkiness of the triphenylsilyl groups, **15** showed high activity (95% yield) in CH_2Cl_2 . However, the resulting **2a** was racemic (entry 4). Very surprisingly, the reaction conducted with **15** (100 mol.%) in toluene gave a 6:4 mixture of **2a** and **3a**, and the subsequent treatment of the mixture with ClSO_3H gave (+)-**2a** with a remarkably high enantiomeric excess (95% enantiomeric excess), although the yield of (+)-**2a** was moderate (57% yield) (entry 5). The absolute configuration of the obtained (+)-**2a** was determined to be (3*S*,5*R*,10*S*) by high performance liquid chromatography analysis after conversion to the known (+)-**11** (ref. 5) by treatment with Bu_3SnH (2 equiv.) and 2,2'-azobisisobutyronitrile (1 mol.%) in toluene at 90°C for 1 h.

Next, the steric effect of the *N*-(1-phenethyl) moiety of the catalyst on the reactivity and enantioselectivity was examined. Interestingly, chiral phosphoramidite **16** bearing an (*R*)-*N*-(1-phenethyl) group hardly dissolved in toluene and was inert for the halocyclization (entry 6). Chiral phosphoramidite **17** bearing an *N*-benzyl group showed low activity (34% yield) and low enantioselectivity (34% enantiomeric excess) (entry 7). Therefore, an (*S*)-*N*-(1-phenethyl) group and a (*R*)-binaphthyl moiety were a matched pair.

Table 2 | Enantioselective halocyclization of 1, 4–6 induced by chiral nucleophilic promoters 13–18.

Entry	Substrate	NXS	Promoter (mol.%)	Solvent	Product	Yield (%)	Enantiomeric excess (%)
1*	1	NIS	13 (30)	CH_2Cl_2	2a	95	0
2	1	NIS	13 (100)	Toluene	2a	11	0
3*	1	NIS	14 (30)	CH_2Cl_2	2a	46	0
4	1	NIS	15 (100)	CH_2Cl_2	2a	95	0
5	1	NIS	15 (100)	Toluene	2a	57	95 (3 <i>S</i>)
6	1	NIS	16 (100)	Toluene	2a	0	not determined
7	1	NIS	17 (100)	Toluene	2a	34	34 (3 <i>S</i>)
8	4	NIS	15 (100)	Toluene	7 + 8	64†	91 (for 7)‡
9	5	NIS	15 (100)	Toluene	9	58	91 (3 <i>S</i>)
10	6	NIS	15 (100)	Toluene	10	52§	99
11	1	NBS	15 (20)	Toluene	2b	29	4‡
12	1	NBS	18 (100)	Toluene	2b	75	36‡

Unless otherwise noted, the reaction of **1**, **4**–**6** (0.1 mmol) and NIS or NBS (1.1 equivalents) in solvent (1 ml) was conducted in the presence of a chiral promoter at -40°C for 24 h. The conversion yield of the product was determined by ^1H nuclear magnetic resonance analysis after treatment with ClSO_3H (50 μl) in 2-nitropropane (2 ml) at -78°C for 4 h. The enantiomeric excess of the product was determined by high performance liquid chromatography analysis.

* The reaction was conducted at -78°C .

† Yield of a 2:1 mixture of **7** and **8**.

‡ The absolute configuration was not determined.

§ Yield of a diastereomeric mixture (94:6 diastereomeric ratio).

To explore the generality and scope of the enantioselective halopolycyclization promoted by **15**, we examined the reaction of 3-(homogeranyl)toluene (**4**) and 4-(homogeranyl)anisole (**5**). These reactions were conducted in toluene at -40°C . On subsequent treatment with ClSO_3H , halogenated monocyclic products in the resultant crude mixtures (the ratio of AB-ring products and A-ring products was $\sim 6:4$) also cyclized to the desired *trans*-fused polycyclic products. The reaction of **4** gave a 2:1 mixture of **7** and **8** in 64% yield (entry 8). Both **7** and **8** were obtained as single *trans*-diastereomers. The enantiomeric excess of the major product **7** was 91%. The reaction of **5** also gave the desired product **9** as a single *trans*-diastereomer in 58% yield with 91% enantiomeric excess (entry 9). The absolute configuration of the obtained **9** was determined to be (3*S*,5*R*,10*S*) by the same procedure as with **2a** after conversion to the known **12** (ref. 15). Furthermore, the reaction of 4-(homofarnesyl)toluene (**6**) gave the desired iodinated tetracyclic compound **10** as a major product along with a byproduct (94:6), with almost complete enantioselectivity (99% enantiomeric excess) in 52% yield (entry 10).

Enantioselective bromo- and chlorocyclizations are more valuable because there are many natural products containing bromo and/or chloro substituents. But unfortunately, the bromocyclization reaction of **1** using **18** gave desired product **2b** with 36% enantiomeric excess in 75% yield (entry 12), and the chlorocyclization of **1** with *N*-chlorosuccinimide by the present method did not give chlorinated product **2c** ($\text{X} = \text{Cl}$). The present methodology was also applicable for enantioselective bromocyclization in principle, although the enantioselectivity was moderate in our present result. For the chemical synthesis of **2b** and **2c** with high enantiomeric excess, transhalogenation of **2a** to **2b** and **2c** was quite effective. Compound **2b** was stereospecifically obtained by lithiation (*t*-BuLi, 3.0 equiv.) of **2a** followed by bromination ($\text{BrCF}_2\text{CF}_2\text{Br}$, 3.0 equiv.) in 85% yield.

Compound **2c** was also obtained using $\text{ClCF}_2\text{CF}_2\text{Cl}$ instead of $\text{BrCF}_2\text{CF}_2\text{Br}$ in 70% yield.

The solvent effect in the halocyclization might be ascribed to the nature of the phosphonium salt generated from **15** and NIS (Fig. 2). In CH_2Cl_2 , the phosphonium salt would be a loose ionic species (the more active species **A**) and the highly ionic nature might result in the high activity of **15**. The free rotation of the *N*-(1-phenethyl)amino group on the P–N axis might lead to no enantioselectivity in CH_2Cl_2 . On the other hand, in a less polar toluene, the phosphonium salt would be a tight ion pair (the less active species **B**) in which the anion of succinimide might form a hydrogen bond with the proton of an NH group to fix the rotation of the *N*-(1-phenethyl)amino group. The less ionic nature might result in the low activity of **15** and the hindered rotation might lead to high enantioselectivity in toluene. When the reaction of **1** with NIS was conducted in toluene in the presence of one equivalent of *N,N*-dimethylformamide (DMF) or 1-ethyl-3-methylimidazolium triflate, the yield of the desired product **2a** was increased but the enantioselectivity was significantly decreased (82% yield, 7% enantiomeric excess and 84% yield, 0% enantiomeric excess, respectively). These polar additives should preferentially interact with the phosphonium salt moiety ($\text{I-P}^+-\text{NH}$) in the active species. This interaction caused a phosphonium salt to lose ionic species **A** in toluene, resulting in high reactivity and poor enantioselectivity for this reaction.

We suggest the following mechanism to explain the absolute stereopreference we observed. **C** and **D** in Fig. 2 are the Newman projections of the ionic species **B** viewed along the I–P bond, and **D** shows only the triphenylsilyl groups and the *N*-(1-phenethyl)amino group of the ion-pair species **B** for clarity. As shown in **D**, space *a* in the upper right is the least hindered and space *c* in the upper left is the most hindered. Therefore, the *si*-face of the terminal isobutenyl

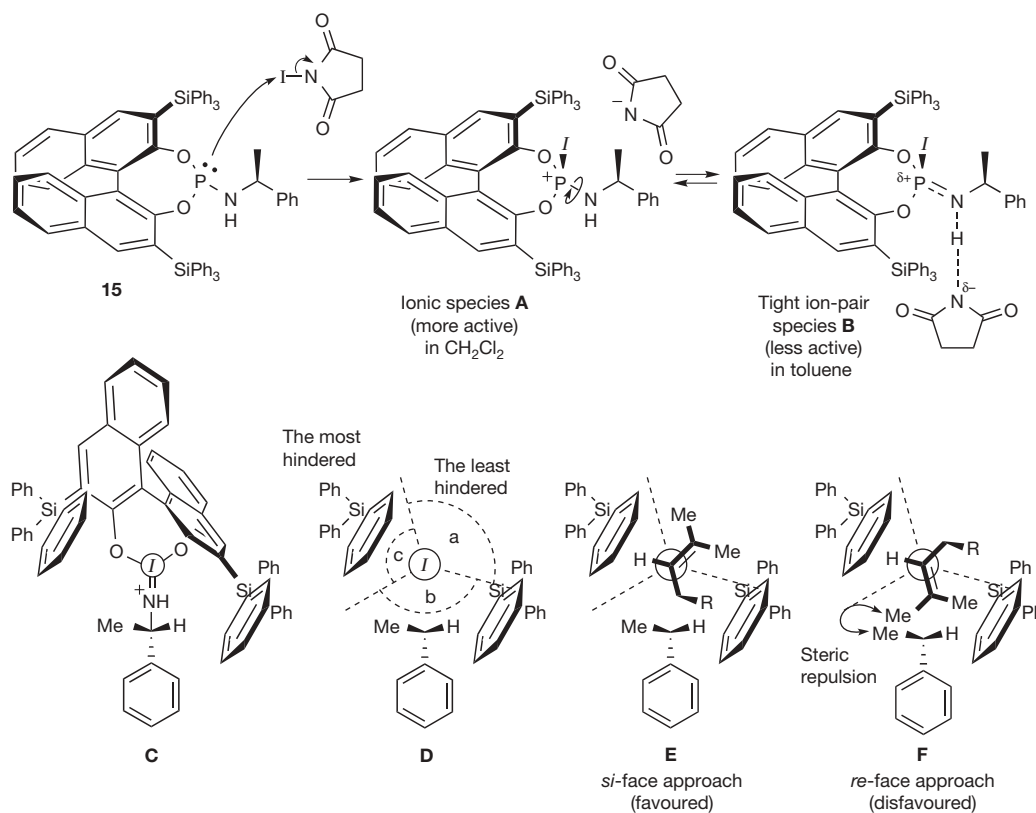


Figure 2 | Possible reaction mechanism. *I* represents activated iodine. In CH_2Cl_2 , the phosphonium salt would be a loose ionic species (the more active species **A**), while a tight ion-pair (the less active species **B**) would be formed in less polar toluene. **C** and **D** are the Newman projections of the ion-pair species **B** viewed along the I–P bond. **E** and **F** represent transition-state

assemblies. **D**, **E** and **F** show only the triphenylsilyl groups and the *N*-(1-phenethyl) group of the ion-pair species **B** for clarity. The *si*-face of the terminal isobutenyl group of the substrates would preferentially approach the activated iodine (**E**).

group of the substrates would preferentially approach the activated iodine (E), while the reaction at the *re*-face would be disfavoured (F). Detailed mechanistic studies are now in progress to confirm this speculative mechanism.

In conclusion, we have developed a method for the halocyclization of homo(polyprenyl)arenes using nucleophilic promoters. This method provides access to the polycyclic 3-iodoterpenoids with high enantioselectivities much more easily than the conventional multi-step syntheses, although enantioselective bromocyclization and chlorocyclization did not work well. Because polycyclic 3-iodoterpenoids can be converted to other halogenated derivatives by stereospecific transhalogenation, the present reaction would be useful for the synthesis of chiral building blocks for organic syntheses. We hope that the results presented here stimulate the development of the method for the enantioselective halocyclizations, and of efficient enantioselective synthesis of the halogenated bioactive natural products.

METHODS

The syntheses of homo(polyprenyl)arenes and chiral phosphoramidites, the procedures of halocyclization and transhalogenation, and the characterization of all new compounds are described in the Supplementary Information.

General procedure for the halocyclization of 1 by nucleophilic phosphorus compounds. To a solution of a nucleophilic phosphorus compound (0.1 mmol) in toluene (1 ml) we added **1** (0.1 mmol) and NIS (0.11 mmol) successively at -78°C , and the mixture was stirred at -78°C for 0.5 h and then at -40°C for 24 h. The reaction mixture was quenched with 20% aqueous $\text{Na}_2\text{S}_2\text{O}_3$ (3 ml) and extracted with hexane (5 ml \times 3). The combined organic layer was dried over MgSO_4 , filtered and concentrated. The residue was purified by column chromatography on silica gel using hexane as an eluent. The resulting mixture of **2a**, endo-**3a** and exo-**3a** was used for the next cyclization without further separation. To a solution of the resulting mixture of **2a**, endo-**3a** and exo-**3a**, which were obtained in the above reaction, in *i*-PrNO₂ (2 ml) we added dropwise ClSO_3H (50 μl) at -78°C . After stirring for 4 h, the reaction mixture was quenched with saturated aqueous NaHCO_3 , and extracted with hexane (5 ml \times 3). The combined organic layer was dried over MgSO_4 , filtered and concentrated. The residue was purified by column chromatography on silica gel using hexane as an eluent, to give **2a**.

Received 26 July; accepted 22 December 2006.

- Blunt, J. W., Copp, B. R., Munro, M. H. G., Northcote, P. T. & Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **23**, 26–78 (2006).
- Yarnell, A. Nature's X-factors. *Chem. Eng. News* **84**, 12–18 (2006).
- Butler, A. & Carter-Franklin, J. N. The role of vanadium bromoperoxidase in the biosynthesis of halogenated marine natural products. *Nat. Prod. Rep.* **21**, 180–188 (2004).

- Yamamura, S. & Terada, Y. Isoaplysin-20, a natural bromine-containing diterpene, from *Aplysia kurodai*. *Tetrahedr. Lett.* **18**, 2171–2172 (1977).
- Ishibashi, H., Ishihara, K. & Yamamoto, H. A new artificial cyclase for polyprenoids: enantioselective total synthesis of (–)-chromazonanol, (+)-8-epi-puuphedione, and (–)-11'-deoxytaondiol methyl ether. *J. Am. Chem. Soc.* **126**, 11122–11123 (2004).
- Yoder, R. A. & Johnston, J. N. A case study in biomimetic total synthesis: polyolefin carbocyclizations to terpenes and steroids. *Chem. Rev.* **105**, 4730–4756 (2005).
- Johnson, W. S. Biomimetic polyene cyclizations: a review. *Bioorg. Chem.* **5**, 51–98 (1976).
- Huang, A. X., Xiong, Z. & Corey, E. J. An exceptionally short and simple enantioselective total synthesis of pentacyclic triterpenes of the β -amyryn family. *J. Am. Chem. Soc.* **121**, 9999–10003 (1999).
- Wolinsky, L. E. & Faulkner, D. J. A biomimetic approach to the synthesis of *Laurencia* metabolites. Synthesis of 10-bromo- α -chamibrene. *J. Org. Chem.* **41**, 597–600 (1976).
- González, A. G., Martin, J. D., Pérez, C. & Ramirez, M. A. Bromonium ion-induced cyclization of methyl farnesate: application to the synthesis of snyderol. *Tetrahedr. Lett.* **17**, 137–138 (1976).
- Yamaguchi, Y., Uyehara, T. & Kato, T. Biogenetic type synthesis of (±)-concinndiol and (±)-aplysin 20. *Tetrahedr. Lett.* **26**, 343–346 (1985).
- Carter-Franklin, J. N., Parrish, J. D., Tschirret-Guth, R. A., Little, R. D. & Butler, A. Vanadium haloperoxidase-catalyzed bromination and cyclization of terpenes. *J. Am. Chem. Soc.* **125**, 3688–3689 (2003).
- Kang, S. H., Lee, S. B. & Park, C. M. Catalytic enantioselective iodocyclization of γ -hydroxy-*cis*-alkenes. *J. Am. Chem. Soc.* **125**, 15748–15749 (2003).
- Zhang, Y., Shibamoto, K. & Yamamoto, H. Lewis-acid mediated selective chlorinations of silyl enolate. *J. Am. Chem. Soc.* **126**, 15038–15039 (2004).
- Grossman, R. B. & Trupp, R. J. The first reagent-controlled asymmetric halolactonizations. Dihydroquinidine-halogen complexes as chiral sources of positive halogen ion. *Can. J. Chem.* **76**, 1233–1237 (1998).
- Cui, X.-L. & Brown, R. S. Mechanistic evaluation of the halocyclization of 4-penten-1-ol by some bis(2-substituted pyridine) and bis(2,6-disubstituted pyridine)bromonium triflates. *J. Org. Chem.* **65**, 5653–5658 (2000).
- Haas, J., Piguel, S. & Wirth, T. Reagent-controlled stereoselective iodolactonizations. *Org. Lett.* **4**, 297–300 (2002).
- Fukuzumi, T., Shibata, N., Sugiura, M., Nakamura, S. & Toru, T. Enantioselective fluorination mediated by cinchona alkaloids/selectfluor combinations: a catalytic approach. *J. Fluor. Chem.* **127**, 548–551 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support for this project was provided by the JSPS (KAKENHI), the 21st Century COE Program "Nature-Guided Materials Processing" of MEXT, the Banyu Award in Synthetic Organic Chemistry, Japan, and the Suzuken Memorial Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.I. (ishihara@cc.nagoya-u.ac.jp).

LETTERS

Large subglacial lakes in East Antarctica at the onset of fast-flowing ice streams

Robin E. Bell¹, Michael Studinger¹, Christopher A. Shuman², Mark A. Fahnestock³ & Ian Joughin⁴

Water plays a crucial role in ice-sheet stability and the onset of ice streams. Subglacial lake water moves between lakes¹ and rapidly drains, causing catastrophic floods². The exact mechanisms by which subglacial lakes influence ice-sheet dynamics are unknown, however, and large subglacial lakes^{3,4} have not been closely associated with rapidly flowing ice streams. Here we use satellite imagery and ice-surface elevations to identify a region of subglacial lakes, similar in total area to Lake Vostok, at the onset region of the Recovery Glacier ice stream in East Antarctica and predicted by ice-sheet models⁵. We define four lakes through extensive, flat, featureless regions of ice surface bounded by upstream troughs and downstream ridges. Using ice velocities determined using interferometric synthetic aperture radar (InSAR), we find the onset of rapid flow (moving at 20 to 30 m yr⁻¹) of the tributaries to the Recovery Glacier ice stream in a 280-km-wide segment at the downslope margins of these four subglacial lakes. We conclude

that the subglacial lakes initiate and maintain rapid ice flow through either active modification of the basal thermal regime of the ice sheet by lake accretion or through scouring bedrock channels in periodic drainage events. We suggest that the role of subglacial lakes needs to be considered in ice-sheet mass balance assessments.

Ice streams are huge fast-flowing features within continental ice sheets that transport inland ice to the grounding line where it is discharged to the ocean, influencing global sea level. In West Antarctica and in Greenland, water and sediments provide basal lubrication to the onset of fast flow of ice streams⁶⁻⁸. The movement of subglacial water between subglacial lakes and along ice streams^{1,9} has been documented through surface elevation changes. In the past 15 million years, large subglacial lakes have repeatedly drained from beneath the East Antarctic ice sheet². Because large subglacial lakes (>1,000 km²) are capable of both substantial modification of the

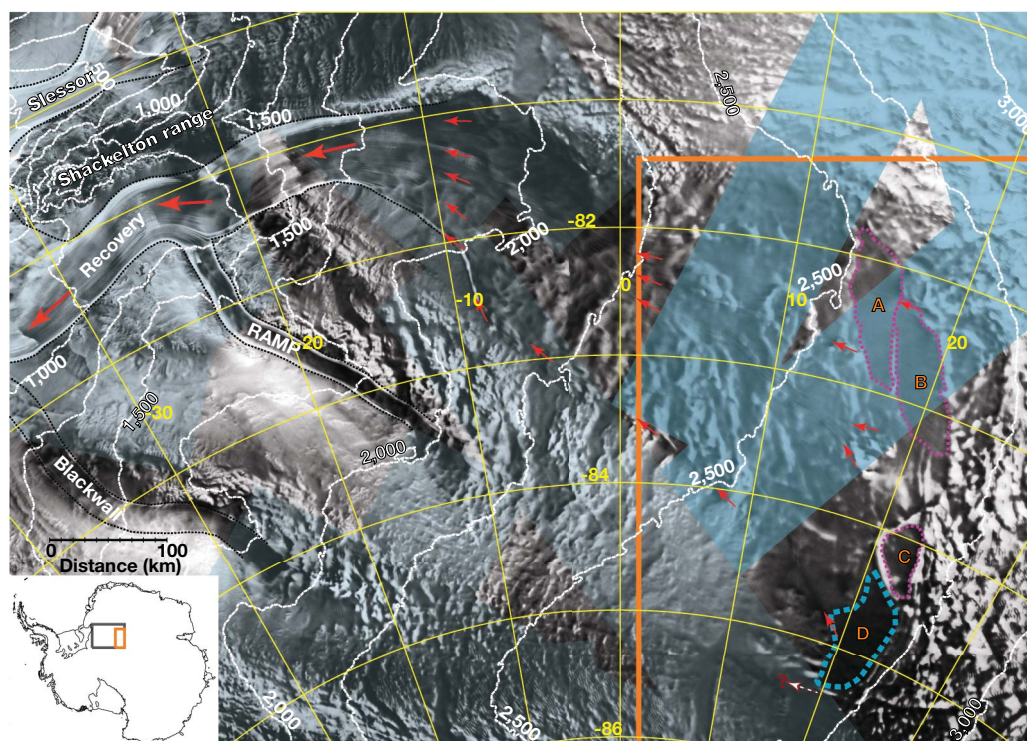


Figure 1 | RADARSAT image of the Recovery ice stream catchment with 250 m ice-surface contours (white) on the basis of the ICESat-derived digital elevation model. Red arrows indicate the location of clearly defined flow features. Blue shading denotes InSAR coverage: light blue, published

velocities¹¹; darker blue, new InSAR velocities. Areas A, B, C and D (dashed blue and purple outlines) are the flat features identified as the Recovery subglacial lakes. The orange box indicates the region detailed in Fig. 2. The inset shows the location of the Recovery ice stream catchment in Antarctica.

¹Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964-8000, USA. ²NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

³Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire 03824, USA. ⁴Applied Physics Lab, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698, USA.

thermal regime of the ice sheet and catastrophic drainage, these lakes have the potential to control the onset of ice streams. Located within 200 km of ice divides, the known large East Antarctic subglacial lakes—Vostok, 90° E, and Sovetskaya—are static, and isolated from the onset of rapid ice flow and the influence of global climate change^{3,4}. A large subglacial lake has been predicted⁵ in the onset region of the Recovery ice stream (this was formerly called the Recovery Glacier ice stream), and bright radar basal reflectors in the onset region of East Antarctic rapid flow are interpreted as subglacial lakes¹⁰. No previous study has clearly linked the onset of rapid ice flow to large subglacial lakes that have sufficient thermal inertia either to modify the ice-sheet thermal regime or to cause outburst flooding.

The Recovery ice stream drains a distinctive funnel-shaped $1 \times 10^6 \text{ km}^2$ catchment that compromises 8% of the East Antarctic ice sheet and contributes 58% of the flux into the Filchner Ice Shelf^{6,11} (Fig. 1). With velocities of 100 m yr^{-1} over 500 km inland⁶, the Recovery ice stream penetrates further into the East Antarctic ice sheet than any other ice stream^{12,13}. In contrast to most ice streams, which get narrower upslope, this ice stream expands from a 30-km-wide feature to a 90-km-wide region east of 15° W. The flow stripes within the Recovery ice stream originate in a 500-km-long region of very flat featureless ice ($<0.45 \text{ m km}^{-1}$) at a surface elevation of 2,540–2,680 m, coincident with a catchment-wide inflection in the

ice-surface slope (Fig. 1). The ice-surface slope upstream and to the east of the flat features is $>2.0 \text{ m km}^{-1}$, while downstream and to the west, the ice-surface slope is $<0.6 \text{ m km}^{-1}$.

High-resolution altimetry data combined with spatial imagery were used to define the horizontal extent of subglacial water bodies^{3,4,14}. The Vostok, 90° E, Concordia and Sovetskaya subglacial lakes are all characterized by a very flat ice surface ($<0.3 \text{ m km}^{-1}$) over the lakes, bounded by troughs 2–15 m deep in the ice surface on the upstream side and ice-surface ridges of 2–5 m on the downstream side. These ice-surface troughs and ridges result from the changing basal stress conditions associated with the transition to a floating ice sheet on the upstream side and the subsequent grounding of the ice sheet on the downstream side¹⁵.

The Recovery ice stream catchment in Queen Maud Land, East Antarctica, remains one of the least explored regions of our planet: it is beyond the coverage of many satellites and was last visited during the 1964–66 surface traverse¹⁶. We used InSAR ice velocities derived from RADARSAT coverage in 1997 and 2000, the Moderate Resolution Imaging Spectroradiometer (MODIS), RADARSAT Antarctic Mapping Program (RAMP) imagery, unpublished ice-penetrating radar data from the 1964–66 Queen Maud Land surface traverse and ice-surface elevations from the Ice, Cloud and land Elevation Satellite (ICESat) to define four structures (A, B, C and D) with the characteristic ice-surface morphology of large subglacial lakes (Fig. 2).

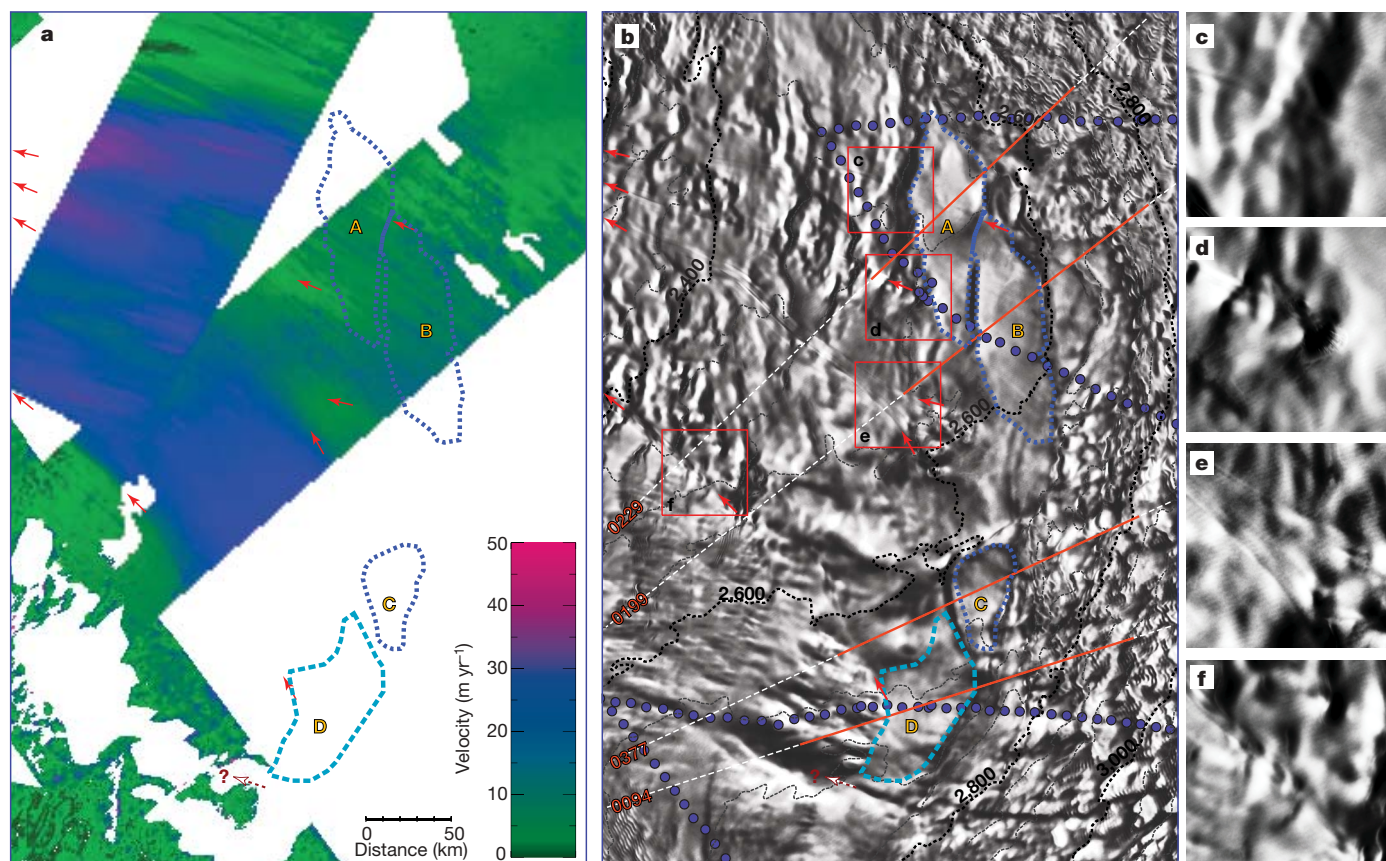


Figure 2 | InSAR ice velocity data and surface imagery of the Recovery ice stream catchment. The location of panels **a** and **b** is presented in Fig. 1. **a**, InSAR relative ice-surface velocity in m yr^{-1} . The southern margin of lake D is close to a distinct topographic trough marked by an arrow and question mark that may indicate a subglacial drainage pathway. **b**, MODIS surface image of the Recovery ice stream catchment with the location of the Recovery subglacial lakes A–D. Included are ICESat-derived surface elevation contours (50 m), the location of clearly defined flow stripes (red arrows) and the route of the 1964–66 South Pole Queen Maud Land traverse (blue dots). The locations of the four ICESat profiles presented in Fig. 3 are shown in dark red and by track number (0229, 0199, 0377 and 0094); the

remainder of their profiles (not shown) are white dashed lines. **c–f**, Enhanced resolution MODIS imagery illustrating flow features over the Recovery subglacial lakes region. The locations of the $50 \text{ km} \times 50 \text{ km}$ insets are shown as red boxes in **b**. **c**, Flow stripe emanating from lake A. **d**, Arcuate crevasse field and associated flow stripes downslope of lake A. **e**, Circular region of rough surface topography between lakes B and C and a 20-km-wide band of flow stripes downstream that can be clearly traced to the middle of the Recovery ice stream. **f**, Linear flow stripe zone coincident with the southern margin of the fast ice flow that can be traced upslope for more than 120 km to the linear surface depressions originating at the downstream margin of lake D.

ICESat data over these features reveals distinctive upstream troughs and downstream ridges. Over four features (A, B, C and D), the ICESat altimetry data and the MODIS imagery shows the upstream troughs to be 3–15 m deep and the downstream ridges to be 2–10 m high (Fig. 3) (with the exception of D, which lacks a well-defined downstream ridge). The surface roughness and the shoreline morphology of these four features closely resemble the known large subglacial lakes such as Vostok and 90° E (Fig. 3). Lake D appears to be a drained lake because the MODIS surface is less homogeneous, and the downstream shoreline ridge is poorly developed. Each of these new lakes is among the largest identified subglacial lakes (A, 3,915 km²; B, 4,385 km²; C, 1,490 km²; D, 3,540 km²) and is comparable in area to lakes 90° E (2,420 km²), and Sovetskaya (1,745 km²). Together, the Recovery subglacial lakes' area (13,300 km²) is similar in scale to that of Lake Vostok (15,690 km²). The ice thickness over these subglacial lakes, from radar and seismic soundings¹⁶, ranges from 3,500 m over A and B to 3,100 m over D (Supplementary Fig. 1). Flexural modelling of the traverse gravity data indicates that the Recovery subglacial lakes are coincident with a tectonic boundary similar to those of lakes Vostok and 90° E but are shallower (Supplementary Fig. 2).

New InSAR velocities¹¹ for the upper reaches of the Recovery ice stream define a 280-km-wide region of elevated ice velocities

(20–30 m yr⁻¹)—the onset of rapid ice flow—that develops along the downstream margin of the Recovery subglacial lakes. This expansive onset region is characterized by crevasses and flow stripes (Fig. 2c) and extends from the northern edge of lake A to the distinctive lineations (Fig. 2f) aligned with the western margin of lake D in the south (Fig. 2b). The southern margin of lake D is also close to a distinct topographic trough indicative of a subsurface drainage pathway (Fig. 2a). Also indicative of the onset of rapid ice flow is the 15-km-wide crevasse field at the downstream margin of lake A, associated with a bedrock peak that rises 1,800 m above the lake surface, as resolved in the traverse radar data (Supplementary Fig. 1)¹⁶. The large crevasse field was encountered by the 1964–66 traverse party and resolved in the MODIS imagery (Fig. 2d). Similarly, east of lake B a circular region of rough surface topography produces a 20-km-wide band of flow stripes downstream that can be clearly traced to the middle of the Recovery ice stream (Fig. 2e). The relatively slow InSAR ice-surface velocities (<5 m yr⁻¹) over these features and downstream flow stripes indicate that these are topographically controlled pinning points within this broad onset region. From our interpretation of the InSAR velocities, flow stripes, topographic slope changes, and other surface features, we conclude that the Recovery subglacial lakes are the onset region of the East Antarctic ice stream, the Recovery ice stream.

This first linkage of large subglacial lakes to the onset of rapid ice flow in the Recovery subglacial lakes region is both directly visible and pervasive. The gently sloping, very broad (280 km) onset region of rapid ice flow is aligned along the full extent of the margins of the Recovery subglacial lakes. These lakes are of fundamental importance to the onset of rapid flow. Earlier studies^{9,10} have suggested that lakes provide a reservoir of water delivering a steady supply of lubricant to the ice stream bed downstream, facilitating the process of rapid sliding and buffering the ice stream base against local freeze-on and associated slowing. If lakes trigger the onset of rapid flow through the continuous delivery of subglacial water, the water drainage would occur through the hydrologic minima, such as relatively narrow valley-like features¹⁷, which would imply that the onset region should be similarly narrow (10–20 km). Here, the width of the onset region (280 km), and the clear spatial linkage between the large subglacial lakes and flow onset, requires other mechanisms to initiate broad accelerated ice flow, such as the modification of the ice-sheet basal thermal regime and periodic, catastrophic drainage.

Large subglacial lakes have the potential to affect the basal thermal conditions of the ice sheet along their full length through the freezing of lake water to the base of the ice sheet, a process observed over large subglacial lakes^{18,19}. As an ice sheet encounters a subglacial lake, basal shear stress drops to zero, resulting in rapid acceleration of the ice sheet. This acceleration will thin the ice sheet, causing the basal thermal gradient to steepen and increasing the rate of heat conduction. For subglacial lakes to cause enhanced flow downstream, the lake must significantly modify this steep basal thermal gradient. The freezing of lake water onto the ice sheet will flatten the basal thermal gradient and reduce the upward flow of heat into the ice²⁰. The lake water acts as a thermal source, warming the basal ice by the heat released during the accretion process for 3,000–70,000 years as the ice above traverses the basin. When this flattened basal thermal gradient is preserved downstream of a lake, the ice sheet will not freeze to the bed when it regrounds downstream and hence streaming flow will result over a broad region.

Water in the largest subglacial lakes may also play a part in regional and global climate in a more catastrophic manner. Sudden lake water releases or outburst floods are known to originate in the interior of East Antarctica and to reach the coast. The Recovery subglacial lakes collect basal water from a larger area (387,890 km²) than any other subglacial lake system yet studied, and subsequently may fill faster and be more dynamic than lakes closer to the ice divides. Assuming a basal melt rate of 1 mm yr⁻¹ over the catchment, the Recovery subglacial lakes have sufficient water input to have filled since the Last

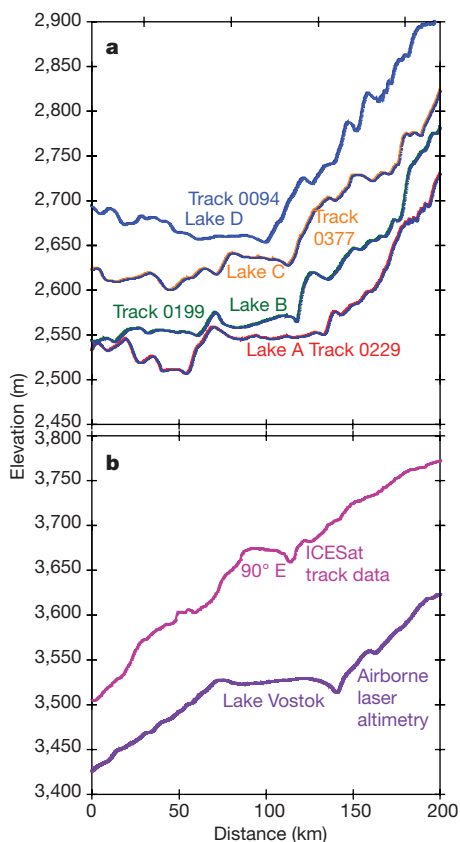


Figure 3 | Detailed elevation profiles across the Recovery subglacial lakes and known lakes. **a**, Recovery subglacial lakes profiles along the ICESat track segments shown in Fig. 2b in dark red. Ice surface morphology characteristic of subglacial lakes, upstream troughs and downstream ridges bounding relatively flat featureless regions is shown. These profiles illustrate the location of Recovery subglacial lakes at major inflections in ice-surface slope between the relatively steep ice-surface slope (>2 m km⁻¹) upstream to the relatively flat ice-surface slope downstream (<0.6 m km⁻¹). Ice-surface elevation over the lakes ranges from 2,540–2,660 m. **b**, Laser altimetry profiles across Lake Vostok (airborne laser altimetry^{4,22}) and the 90° E lake (ICESat). The ice-surface slope on both sides of these lakes is >1.4 m km⁻¹. Ice-surface elevation over these lakes ranges from 3,500 to 3,670 m, and all elevations have been geoid-corrected.

Glacial Maximum, unlike Vostok and other large interior lakes. Rapidly filling lakes may periodically drain, flush down the ice streams and scour bedrock channels, enhancing rapid ice flow the same way that basal water produces rapid ice motion during a glacier surge²¹. An outburst similar in scale to the lake-to-lake drainage in East Antarctica¹ (1.8 km³) would have a periodicity of ~4.5 years, whereas an event of scale of the order of the East Antarctic outburst floods² would have a periodicity of thousands of years. Episodic draining of large subglacial lakes provides another mechanism for the initiation of climatically significant rapid ice flow with ice streams developing along the scoured channels.

The Recovery subglacial lakes capture water from a large area, effectively concentrating the energy from basal melting and re-releasing it where it can have a significant impact on ice flow through either basal accretion or catastrophic drainage. Contributing 35 gigatons per year of ice to the global oceans¹¹, the Recovery subglacial lakes and the associated Recovery ice stream tributaries have the potential greatly to affect the drainage of the East Antarctic ice sheet and its influence on sea level rise in the near future. Subglacial lakes and the associated hydrologic systems are crucial components in the dynamic evolution of ice sheets and need to be incorporated into ice-sheet models that are used for climate predictions.

Received 26 June; accepted 27 December 2006.

- Wingham, D. J., Siegert, M. J., Shepherd, A. & Muir, A. S. Rapid discharge connects Antarctic subglacial lakes. *Nature* **440**, 1033–1036 (2006).
- Lewis, A. R., Marchant, D. R., Kowalewski, D. E., Baldwin, S. L. & Webb, L. E. The age and origin of the Labyrinth, western Dry Valleys, Antarctica: Evidence for extensive middle Miocene subglacial floods and freshwater discharge to the Southern Ocean. *Geology* **34**, 513–516 (2006).
- Kapitsa, A. P., Ridley, J. K., Robin, G. D., Siegert, M. J. & Zotikov, I. A. A large deep freshwater lake beneath the ice of central East Antarctica. *Nature* **381**, 684–686 (1996).
- Bell, R. E., Studinger, M., Fahnestock, M. & Shuman, C. A. Tectonically controlled subglacial lakes on the flanks of the Gamburtsev Subglacial Mountains, East Antarctica. *Geophys. Res. Lett.* **33**, L02504 (2006).
- Johnson, J. V. A *Basal Model for Ice Sheets*. PhD thesis, Univ. Maine (2002).
- Joughin, I., Bamber, J. L., Scambos, T., Tulaczyk, S. & Fahnestock, M. Integrating satellite observations with modelling: basal shear stress of the Filchner-Ronne ice streams, Antarctica. *Phil. Trans. R. Soc. A* **364**, 1795–1814 (2006); doi:10.1098/rsta.2006.1799.
- Fahnestock, M., Abdalati, W., Joughin, I., Brozena, J. & Gogineni, P. High geothermal heat flow, basal melt, and the origin of rapid ice flow in central Greenland. *Science* **294**, 2338–2342 (2001).
- Bell, R. E. *et al.* Influence of subglacial geology on the onset of a West Antarctic ice stream from aerogeophysical observations. *Nature* **394**, 58–62 (1998).
- Gray, L. *et al.* Evidence for subglacial water transport in the West Antarctic Ice Sheet through three-dimensional satellite radar interferometry. *Geophys. Res. Lett.* **32**, L03501, doi:10.1029/2004GL021387 (2005).
- Siegert, M. J. & Bamber, J. L. Subglacial water at the heads of Antarctic ice stream tributaries. *J. Glaciol.* **46**, 702–703 (2000).
- Joughin, I. & Bamber, J. L. Thickening of the ice stream catchments feeding the Filchner-Ronne Ice Shelf, Antarctica. *Geophys. Res. Lett.* **32**, L17503, doi:10.1029/2005GL023844 (2005).
- Jezek, K. C. RADARSAT-1 Antarctic Mapping Project: change-detection and surface velocity campaign. *Ann. Glaciol.* **34**, 263–268 (2002).
- Jezek, K. C., Farness, K., Carande, R., Wu, X. & Labelle-Hamer, N. RADARSAT 1 synthetic aperture radar observations of Antarctica: Modified Antarctic Mapping Mission, 2000. *Radio Sci.* **38**, doi:10.1029/2002RS002643 (2003).
- Cudlip, W. & McIntyre, N. F. SEASAT altimeter observations of an Antarctic “lake”. *Ann. Glaciol.* **9**, 55–59 (1987).
- Gudmundsson, G. H. Transmission of basal variability to a glacier surface. *J. Geophys. Res.* **108**, doi:10.1029/2002JB002107 (2003).
- Beitzel, J. E. in *Antarctic Snow and Ice Studies* Vol. II, 39–87 (American Geophysical Union, Washington DC, 1971).
- Tikku, A. A., Bell, R. E., Studinger, M. & Clarke, G. K. C. Ice flow field over Lake Vostok, East Antarctica inferred by structure tracking. *Earth Planet. Sci. Lett.* **227**, 249–261 (2004).
- Bell, R. E., Studinger, M., Tikku, A. A., Clarke, G. K. C., Gutner, M. M. & Meertens, C. Origin and fate of Lake Vostok water frozen to the base of the East Antarctic ice sheet. *Nature* **416**, 307–310 (2002).
- Jouzel, J. *et al.* More than 200 meters of lake ice above subglacial Lake Vostok. *Science* **286**, 2138–2141 (1999).
- Hulbe, C. L. & Fahnestock, M. A. West Antarctic ice-stream discharge variability: mechanism, controls and pattern of grounding-line retreat. *J. Glaciol.* **50**, 471–484 (2004).
- Kamb, B. *et al.* Glacial surge mechanism 1982–1983 surge of variegated glacier, Alaska. *Science* **227**, 269–279 (1985).
- Studinger, M. *et al.* Ice cover, landscape setting, and geological framework of Lake Vostok, East Antarctica. *Earth Planet. Sci. Lett.* **205**, 195–210 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. K. C. Clarke and R. Bindshadler for comments on an earlier version. A 2003 presentation prepared by G. Leitchenkov for the SCAR SALE meeting stimulated this effort. J. Clough, C. Bentley, L. Kissel of the Byrd Polar Research Center Archival Program and B. R. Bell facilitated the rescue of the radar and seismic data along the Recovery subglacial lakes portion of the 1964–1966 QML Traverse route. Technical assistance with the ICESat, RAMP and MODIS data was provided by V. Suchdeo. R.E.B. and M.S. were supported by the Doherty Endowment of LDEO and the Palisades Geophysical Institute. M.A.F., C.A.S. and I.J. were supported by NASA’s Cryospheric programme. The US National Science Foundation provided support for the fundamental studies of Lake Vostok.

Author Contributions All authors discussed the results and commented on the manuscript. R.E.B. led the development of this paper and the integration of the results, M.S. analysed the surface traverse data and modelled the gravity data, C.A.S. led the ICESat, RAMP and MODIS integration and interpretation, M.A.F. contributed to the visible-band image analysis and the conceptual models of the onset mechanism and I.J. provided the InSAR ice velocities.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.E.B. (robinb@ldeo.columbia.edu).

LETTERS

Moisture transport across Central America as a positive feedback on abrupt climatic changes

Guillaume Leduc¹, Laurence Vidal¹, Kazuyo Tachikawa¹, Frauke Rostek¹, Corinne Sonzogni¹, Luc Beaufort¹ & Edouard Bard¹

Moisture transport from the Atlantic to the Pacific ocean across Central America leads to relatively high salinities in the North Atlantic Ocean¹ and contributes to the formation of North Atlantic Deep Water². This deep water formation varied strongly between Dansgaard/Oeschger interstadials and Heinrich events—millennial-scale abrupt warm and cold events, respectively, during the last glacial period³. Increases in the moisture transport across Central America have been proposed to coincide with northerly shifts of the Intertropical Convergence Zone and with Dansgaard/Oeschger interstadials, with opposite changes for Heinrich events⁴. Here we reconstruct sea surface salinities in the eastern equatorial Pacific Ocean over the past 90,000 years by comparing palaeotemperature estimates from alkenones and Mg/Ca ratios with foraminiferal oxygen isotope ratios that vary with both temperature and salinity. We detect millennial-scale fluctuations of sea surface salinities in the eastern equatorial Pacific Ocean of up to two to four practical salinity units. High salinities are associated with the southward migration of the tropical Atlantic Intertropical Convergence Zone, coinciding with Heinrich events and with Greenland stadials⁵. The amplitudes of these salinity variations are significantly larger on the Pacific side of the Panama isthmus, as inferred from a comparison of our data with a palaeoclimate record from the Caribbean basin⁶. We conclude that millennial-scale fluctuations of moisture transport constitute an important feedback mechanism for abrupt climate changes, modulating the North Atlantic freshwater budget and hence North Atlantic Deep Water formation.

Paleotemperatures recorded in Greenland indicate that rapid climatic changes occurred on a millennial timescale during Marine Isotope Stage 3 (MIS3^{5,7}, between 59 and 25 kyr BP) and were intimately linked to the process of North Atlantic Deep Water (NADW) formation^{2,3} (the Dansgaard–Oeschger⁵ and Heinrich events⁷, H-DO variability). The H-DO features are also observed in tropical areas sensitive to summer monsoon fluctuations, which were linked to latitudinal shifts of the Intertropical Convergence Zone (ITCZ)^{4,8,9}. To explain the existence of H-DO features at different latitudes, we need to understand how climatic teleconnections are set up between high and low latitudes and/or different ocean basins^{10,11}.

In modern climatology, the interoceanic freshwater transfer from the Atlantic to the Pacific Oceans maintains the high Atlantic salinity required for NADW formation¹. This is achieved by the combination of strong easterly winds and increased atmospheric humidity within the Caribbean region¹², which occurs mainly during the boreal summer when the ITCZ is shifted northward (Supplementary Information). Modelling studies imply that the modern interoceanic vapour flux ranges from 0.13 Sv (ref. 1) to 0.45 Sv (ref. 13) (1 Sv = 10⁶ m³ s⁻¹).

For the Last Glacial Maximum (LGM), different models show either an increased¹⁴ or a decreased¹⁵ water vapour export from the

Atlantic Ocean. Recent work on marine sequences covering the last deglaciation suggested that cross-isthmus vapour transport was not strongly affected by glacial boundary conditions¹⁶. Rather, the interoceanic vapour transport and the interbasin salt contrast seem to have been influenced by ITCZ dynamics, which contributed to the hydrological variability in the Eastern Equatorial Pacific (EEP) region¹⁶. From this point of view, the EEP appears to be a key climatic crossroad involved in rapid climatic changes. To understand better how the ITCZ dynamics and the moisture transfer changes across Central America evolved with respect to the H-DO variability, it is crucial to gather observations covering the entire MIS3, and in particular at the prominent six Heinrich events. Since the EEP sea surface salinities (SSS) are intimately linked to the intensity of the ITCZ and its mean latitudinal position (Fig. 1), reconstructing past EEP SSS fluctuations allows us to track variations in moisture transport across the Panama isthmus (Supplementary Information).

Here we present a multi-proxy record of past EEP hydrological characteristics from the sediment core MD02-2529 (08° 12.33' N, 84° 07.32' W, 1,619 m water depth) (Fig. 2). The $\delta^{18}\text{O}_{G. ruber}$ values for the surface-dwelling foraminifer *Globigerinoides ruber* ($\delta^{18}\text{O}_{G. ruber}$) are about -3‰ for the Holocene and decrease by about 2‰ between the MIS2 and the late Holocene (Fig. 2a). Rapid $\delta^{18}\text{O}_{G. ruber}$ variations of 0.5 to 1‰ occurred on a millennial timescale during MIS2 and the late part of MIS3. Two longer cycles with saw-tooth shapes are recorded between 40 and 46 kyr BP and between 46 and 53 kyr BP. From 58 to 90 kyr BP, fluctuations are observed with longer wavelengths and weaker amplitudes than the $\delta^{18}\text{O}_{G. ruber}$ shifts during MIS2 and 3.

The U_{37}^K sea surface temperatures (U_{37}^K -SST) reconstruction agrees well with the record of the Mg/Ca-SST measured on *G. ruber*, except between 20 and 25 kyr BP (Fig. 2b, Supplementary Information). The mean Holocene SST of 28 °C (Fig. 2b) is within the range of modern SST values (Supplementary Information). A long-term 3 °C increase occurred between 25 and 10 kyr BP in U_{37}^K -SST. The SST record for MIS3, 4 and 5 varies between 24.5 and 28 °C with minima and maxima centred at 30, 43 and 65 kyr BP, and at 38, 57 and 85 kyr BP, respectively. Therefore the EEP SST appear to vary independently from the millennial-scale variability of $\delta^{18}\text{O}_{G. ruber}$ (Fig. 2a, b), and the $\delta^{18}\text{O}$ of sea water ($\delta^{18}\text{O}_{sw}$) clearly indicates that the large-amplitude millennial-scale $\delta^{18}\text{O}_{G. ruber}$ fluctuations are driven by $\delta^{18}\text{O}_{sw}$ variations (Fig. 2a, c and Supplementary Information).

After the removal of the global $\delta^{18}\text{O}_{sw}$ linked to sea-level variations (ref. 17), rapid fluctuations in $\Delta\delta^{18}\text{O}_{sw}$ (a proxy for regional SSS) persist throughout the sequence (Fig. 3c). For the modern climate, a northern position of the ITCZ during the boreal summer leads to increased net precipitation in the EEP region (Fig. 1). Indeed, about half of the EEP precipitation originates from the Caribbean¹⁸ and is brought to the Pacific side through zonal atmospheric transport

¹CEREGE, UMR6635, CNRS Université Paul Cézanne Aix-Marseille III, Collège de France, Europôle de l'Arbois, BP 80, 13545 Aix-en-Provence Cedex 04, France.

across Central America, mainly during the wet summer season¹² (Supplementary Information). The salinity record based on $\Delta\delta^{18}\text{O}_{\text{sw}}$ can therefore be used as an indicator of rainfall and, in connection with similar records, of cross-isthmus moisture transport. Time intervals with low $\Delta\delta^{18}\text{O}_{\text{sw}}$ should correspond to periods of intensified moisture fluxes across the Panama isthmus.

The high-resolution reflectance record of the Cariaco Basin (Fig. 1) results from high/low productivity and terrigenous input associated with the northward/southward shifts of the ITCZ over the last glacial period⁴. The northern position of the ITCZ appears to be in phase with DO interstadials⁴, while southward migrations of the ITCZ are in phase with Heinrich events and DO stadials (Fig. 3a, b). The MD02-2529 $\Delta\delta^{18}\text{O}_{\text{sw}}$ and Cariaco Basin reflectance data indicate that the northward migration of the ITCZ in the Cariaco Basin is associated with EEP SSS minima, and so to enhanced moisture export from the Atlantic, when the ITCZ was aligned with the Central America low-level mountain channels⁴ (Fig. 3b, c). Therefore, the Equatorial Atlantic and the EEP are efficiently linked through atmospheric teleconnection.

Maximum EEP SSS occurred during H events and DO stadials (Fig. 3c). This is compatible with a southward migration of the ITCZ accompanied by a decrease in water vapour transport to the Pacific Ocean. The ITCZ displacements are also documented in northeastern Brazil at 10°S by well-dated growth phases of

speleothems during the prominent H events, indicating wet climate at these periods¹⁹ (Fig. 3d). The modern distribution of precipitation over South America in March (Fig. 1) implies a southward migration of the ITCZ over the Amazon Basin at the times of H events (Fig. 3d). These enhanced rainfall time intervals correspond to periods of increased runoff to the tropical Atlantic, as is clearly recorded in shallow sediments along the northeastern Brazilian coast²⁰ (Figs 1 and 3d). The orogenic blocking by the Andes prevented the export of water vapour from the Atlantic to the Pacific, and probably induced the recirculation of freshwater within the Atlantic Ocean during H events and DO stadials, mainly via the Amazon River outflow.

We consider similar records obtained from cores retrieved on the other side of the Panama isthmus in order to better constrain the EEP hydrological changes and to further study the moisture transport from the Atlantic to the Pacific. Recent work on a Caribbean sediment core (Fig. 1) has recorded regional SSS maxima during Heinrich events and the Younger Dryas⁶, indicating that both sides of the isthmus were strongly influenced by north–south shifts of the ITCZ (Fig. 1), making the Atlantic and the Pacific salinity records in phase for their main features. However, a second-order antiphase pattern is superimposed on these first-order variations, and is linked to cross-isthmus moisture transport. Evidence for this second-order antiphase is seen when the respective amplitudes of salinity changes

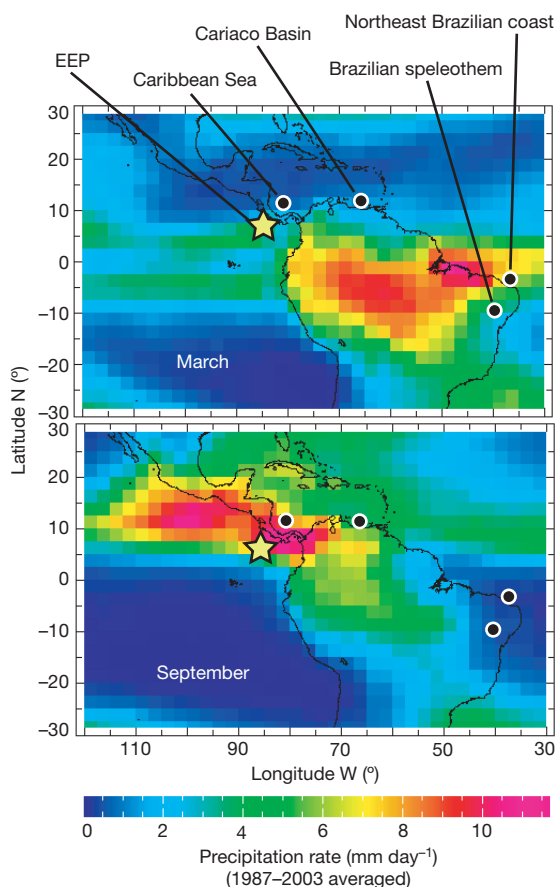


Figure 1 | Averaged precipitation rates over South America for March and September for the period AD 1987–2003. Black dots indicate locations of palaeoclimatic archives discussed in the text, that is, the comparison of the EEP (core MD02-2529, yellow star, this study) to the sedimentary sequences of the Cariaco Basin (ODP hole 1002C)⁴, of the Caribbean Sea (core VM28-122 and ODP hole 999A)⁶ and of the northeastern Brazilian margin (core GeoB3104-1/3912-1)²⁰ as well as the northeastern Brazil speleothem growing phases¹⁹. Rainfall data were retrieved from the International Research Institute for Climate Prediction and are available at <http://iri.ldeo.columbia.edu>.

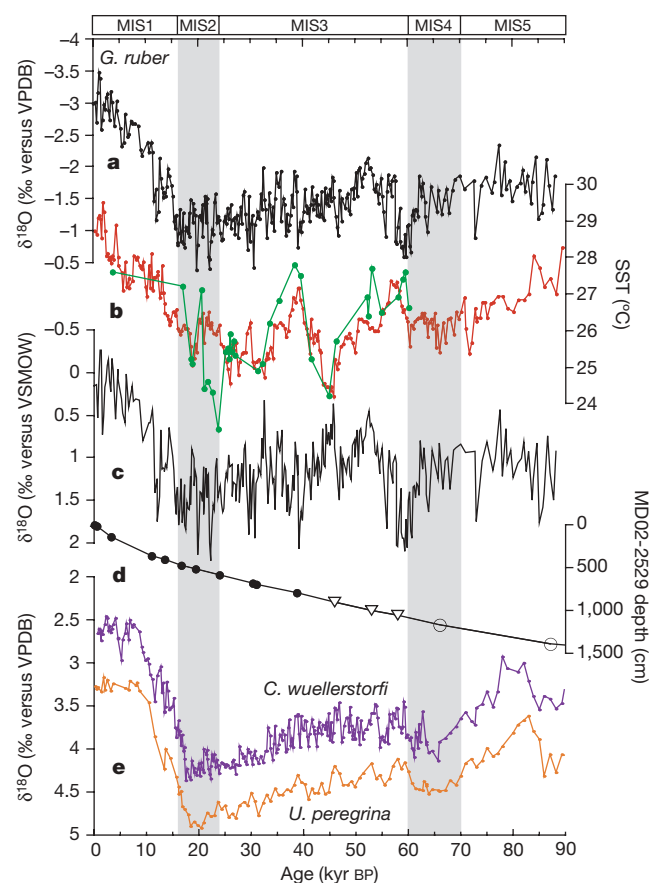


Figure 2 | Records of MD02-2529 versus age for the last 90 kyr BP. **a**, $\delta^{18}\text{O}$ record of the surface-dwelling planktonic foraminifer *G. ruber*. **b**, Comparison of U_{37}^T -SST (red curve) and Mg/Ca-SST (green curve). **c**, Calculated $\delta^{18}\text{O}_{\text{sw}}$. **d**, Age–depth relationship based on radiocarbon measurements (black dots) and benthic foraminifera stratigraphy tuned to Byrd (open triangles) and to a benthic $\delta^{18}\text{O}$ stack (open circles) (see Supplementary Information for the age model construction). **e**, $\delta^{18}\text{O}$ records of the benthic species *Cibicides wuellerstorfi* (blue curve) and *Uvigerina peregrina* (orange curve). VPDB, Vienna Pee-Dee Belemnite standard; VSMOW, Vienna Standard Mean Ocean Water.

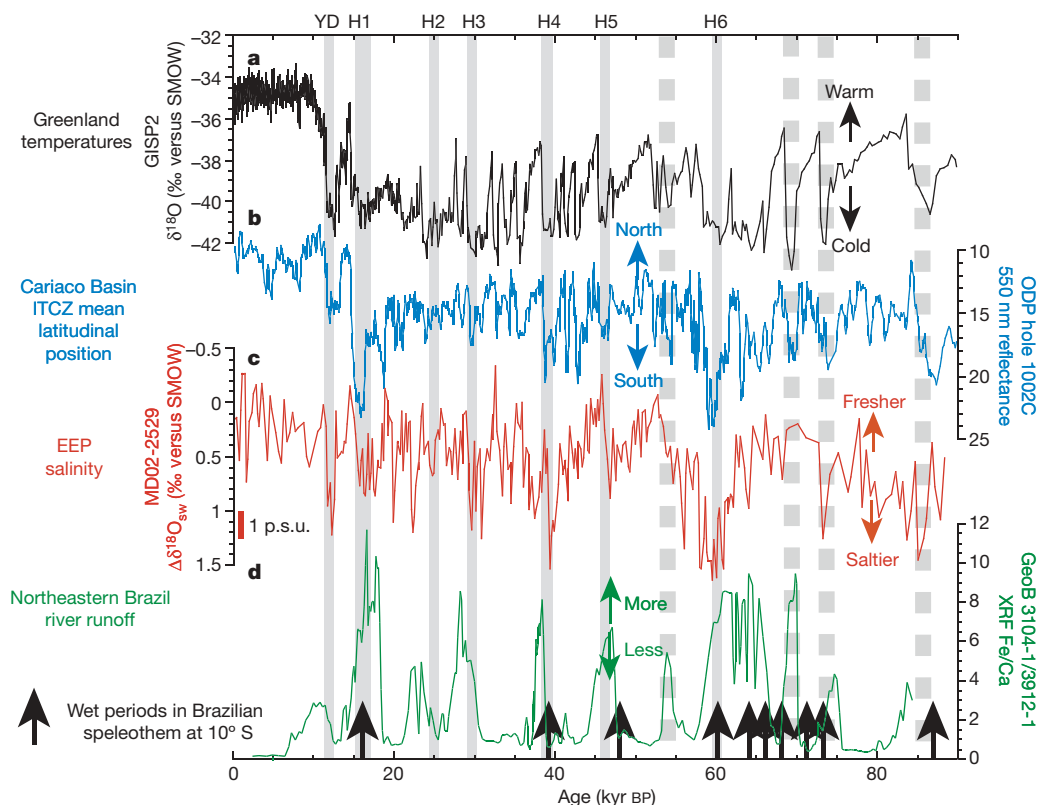


Figure 3 | Temporal variations of the calculated $\Delta\delta^{18}\text{O}_{\text{sw}}$ of MD02-2529 compared to other palaeoclimatic records. The data are presented on their published original timescales without further tuning. **a**, Greenland GISP2 palaeotemperature record⁵. **b**, Cariaco basin sediment reflectance, monitoring the latitudinal mean position of the ITCZ in the northwestern equatorial Atlantic⁴. **c**, $\Delta\delta^{18}\text{O}_{\text{sw}}$ of MD02-2529. **d**, Fe/Ca measurements, a

proxy for riverine input, performed on sediments retrieved off Brazil²⁰; black arrows indicate time intervals of the ITCZ southward expansion over Brazil based on well-dated speleothem growth intervals¹⁹. Grey vertical bars mark the North Atlantic H events. See Fig. 1 for the localization of tropical palaeoclimatic archives. YD, Younger Dryas; H1–H6, Heinrich events.

in the Caribbean and the EEP are considered (Supplementary Information). Our observations demonstrate that for all Heinrich events and the Younger Dryas the salinity increases are about two to three times larger on the Pacific side than on the Atlantic side of the isthmus (Supplementary Information). Our interpretation is that the salinity increases on both sides of the isthmus are modulated by the southward ITCZ shifts, whereas the signature of decreased inter-oceanic moisture flux is less salinity increase on the Atlantic side and more salinity increase on the Pacific side. This is precisely the anti-phased pattern of cross-isthmus moisture transport changes.

The MD02-2529 $\Delta\delta^{18}\text{O}_{\text{sw}}$ values imply that a northerly location of the ITCZ allowed enhanced moisture transport across the Panama isthmus, potentially leading to the build-up of salt in North Atlantic surface waters. By contrast, a southerly ITCZ position led to the orogenic blocking of moisture transport by the Andes⁴. This fresh water returned preferentially into the Atlantic Ocean, in particular through the Amazon basin drainage²⁰. Inevitably, this must have lowered the salinity of low-latitude currents in the Atlantic Ocean, such as the North Brazil current, the Guyana current and the Caribbean current feeding the Gulf Stream.

The deuterium excess record from GRIP ice cores in Greenland shows that changes in their source temperature reflect southward shifts of the geographical location of moisture sources during stadials, presumably associated with southward displacements of the ITCZ²¹. This provides additional evidence for the key role played by the atmospheric water cycle and its related cross-isthmus moisture transport with respect to the H-DO variability. Some modelling studies attempted to reproduce the H-DO variability related to the formation of NADW by applying small freshwater anomalies of 0.03 Sv to the North Atlantic surface water²² that are one order of magnitude smaller than the modern cross-isthmus

vapour transport^{1,13,14,18}. In the modern North Atlantic, low-salinity anomalies initially located at river mouths²³ are then advected in the Nordic Seas after a delay of several years with a probable influence on NADW formation²⁴. If the ITCZ latitudinal cycles were indeed able to modulate the inter-oceanic salt contrast at the millennial timescale, then they could have contributed to the freshwater flux forcing applied in modelling experiments²². It is also possible that further north in the North Atlantic gyre other feedbacks could operate in a different manner²⁵. The combined palaeoceanographic evidence and all identified feedback effects should be considered in modelling these events, that may enlighten our understanding of the modulation of NADW formation process and its related climatic effects.

We speculate that the moisture transport across Central America acts as a positive feedback on abrupt climatic changes. If true, the tropical climate variability is not limited to a passive response to the abrupt climatic changes observed in the North Atlantic region. Further tests of the implications of observed latitudinal and longitudinal hydrological shifts should be performed with models coupling atmosphere and ocean processes at the seasonal scale, and using highly resolved representations of Central and South-America in order to simulate realistic runoff fluxes.

METHODS

We estimated the regional SSS over the last 90 kyr BP by measuring $\delta^{18}\text{O}_{\text{G. ruber}}$ coupled with alkenone-based (using U_{37}^K) and *G. ruber* Mg/Ca-based (using the cleaning method including the reductive step) sea surface palaeotemperatures data (Supplementary Information). To convert the U_{37}^K and the Mg/Ca into temperatures, we have used the calibration of refs 26 and 27, respectively. The lower Mg/Ca temperatures of the 20–25 kyr BP time interval might be caused by foraminiferal test dissolution (Fig. 2b, Supplementary Information). The $\delta^{18}\text{O}$ of sea water ($\delta^{18}\text{O}_{\text{sw}}$) was estimated by combining the $\delta^{18}\text{O}_{\text{G. ruber}}$ and U_{37}^K -SST (Fig. 2c, Supplementary Information). To reconstruct past EEP salinities

($\Delta\delta^{18}\text{O}_{\text{sw}}$, Fig. 3c) we also removed the effect of global ice volume changes (ref. 17) from $\delta^{18}\text{O}_{\text{sw}}$. In the modern EEP, a $\delta^{18}\text{O}_{\text{sw}}$ drop of 1‰ corresponds to a salinity decrease of about 4 p.s.u. (ref. 18), remaining highly significant with respect to the typical error on salinity reconstructions²⁸ (Supplementary Information). The age model is constrained by a series of radiocarbon ages and by isotopic stratigraphy using the $\delta^{18}\text{O}$ temporal variations measured on two benthic foraminifer species (Fig. 2d, e; see Supplementary Information).

Received 18 October 2006; accepted 4 January 2007.

1. Zaucker, F. & Broecker, W. S. The influence of atmospheric moisture transport on the fresh water balance of the Atlantic drainage basin: General circulation model simulations and observations. *J. Geophys. Res.* **97**, 2765–2773 (1992).
2. Broecker, W. S., Bond, G., Klas, M., Bonani, G. & Wolfli, W. A salt oscillator in the glacial Atlantic? 1. The concept. *Paleoceanography* **5**, 469–477 (1990).
3. Bard, E. Climate shock: Abrupt changes over millennial time scales. *Phys. Today* **55**, 32–37 (2002).
4. Peterson, L. C., Haug, G. H., Hughen, K. A. & Röhl, U. Rapid changes in the hydrologic cycle of the tropical Atlantic during the Last Glacial. *Science* **290**, 1947–1951 (2000).
5. Stuiver, M. & Grootes, P. M. GISP2 oxygen isotope ratios. *Quat. Res.* **53**, 277–284 (2000).
6. Schmidt, M. W., Spero, H. J. & Lea, D. W. Links between salinity variation in the Caribbean and North Atlantic thermohaline circulation. *Nature* **428**, 160–163 (2004).
7. Bond, G. *et al.* Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
8. Wang, Y. J. *et al.* A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345–2348 (2001).
9. Ivanochko, T. S. *et al.* Variations in tropical convection as an amplifier of global climate change at the millennial scale. *Earth Planet. Sci. Lett.* **235**, 302–314 (2005).
10. Broecker, W. S. Does the trigger for abrupt climate change reside in the ocean or in the atmosphere? *Science* **300**, 1519–1522 (2003).
11. Vidal, L. & Arz, H. in *Past Climate Variability Through Europe And Africa* (eds Battarbee, R. W. *et al.*) 31–44 (Springer, Dordrecht, 2004).
12. Liu, W. T. & Tang, W. Estimating moisture transport over oceans using space-based observations. *J. Geophys. Res.* **110**, D10101, doi:10.1029/2004JD005300 (2005).
13. Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **1**, 841–866 (1988).
14. Hostetler, S. W. & Mix, A. C. Reassessment of ice-age cooling of the tropical ocean and atmosphere. *Nature* **399**, 673–676 (1999).
15. Schmittner, A., Meissner, K. J., Eby, M. & Weaver, A. J. Forcing of the deep ocean circulation in simulations of the Last Glacial Maximum. *Paleoceanography* **17**, 1015, doi:10.1029/2001PA000633 (2002).
16. Benway, H. M., Mix, A. C., Haley, B. A. & Klinkhammer, G. P. Eastern Pacific Warm Pool paleosalinity and climate variability: 0–30 kyr. *Paleoceanography* **21**, PA3008, doi:10.1029/2005PA001208 (2006).
17. Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).
18. Benway, H. M. & Mix, A. C. Oxygen isotopes, upper-ocean salinity, and precipitation sources in the eastern tropical Pacific. *Earth Planet. Sci. Lett.* **224**, 493–507 (2004).
19. Wang, X. *et al.* Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. *Nature* **432**, 740–743 (2004).
20. Arz, H. W., Pätzold, J. & Wefer, G. Correlated millennial-scale changes in surface hydrography and terrigenous sediment yield inferred from last-glacial marine deposits off northeastern Brazil. *Quat. Res.* **50**, 157–166 (1998).
21. Masson-Delmotte, V. *et al.* GRIP deuterium excess reveals rapid and orbital-scale changes in Greenland moisture origin. *Science* **309**, 118–121 (2005).
22. Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158 (2001).
23. Masson, S. & Delecluse, P. Influence of the Amazon river runoff on the tropical Atlantic. *Phys. Chem. Earth B* **26**, 137–142 (2001).
24. Mignot, J. & Frankignoul, C. Interannual to interdecadal variability of sea surface salinity in the Atlantic and its link to the atmosphere in a coupled model. *J. Geophys. Res.* **109**, C04005, doi:10.1029/2003JC002005 (2004).
25. Schmidt, M. W., Vautravers, M. J. & Spero, H. J. Rapid subtropical North Atlantic salinity oscillations across Dansgaard-Oeschger cycles. *Nature* **443**, 561–564 (2006).
26. Sonzogni, C. *et al.* Temperature and salinity effects on alkenone ratios measured in surface sediments from the Indian Ocean. *Quat. Res.* **47**, 344–355 (1997).
27. Lea, D. W., Pak, D. K. & Spero, H. J. Climate impact of Late Quaternary equatorial Pacific sea surface temperature variations. *Science* **289**, 1719–1724 (2000).
28. Schmidt, G. A. Error analysis of paleosalinity calculations. *Paleoceanography* **14**, 422–429 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from INSU and the French Polar Institute IPEV, which provided the RV *Marion Dufresne* and the CALYPSO coring system used during the IMAGES VIII MONA cruise. Thanks to Y. Garcin and M. Siddall for discussion and reviews. Paleoclimate work at CEREGE is supported by grants from the CNRS, the ANR and the Gary Comer Science and Education Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.L. (leduc@cerege.fr) and E.B. (bard@cerege.fr).

LETTERS

Female fur seals show active choice for males that are heterozygous and unrelated

J. I. Hoffman¹, J. Forcada², P. N. Trathan² & W. Amos¹

Much debate surrounds the exact rules that influence mating behaviour, and in particular the selective forces that explain the evolution of female preferences. A key example is the lek paradox, in which female choice is expected rapidly to become ineffective owing to loss of additive genetic variability for the preferred traits^{1–3}. Here we exploit a remarkable system in which female fur seals exert choice by moving across a crowded breeding colony to visit largely static males. We show that females move further to maximize the balance between male high multilocus heterozygosity and low relatedness. Such a system shows that female choice can be important even in a strongly polygynous species, and at the same time may help to resolve the lek paradox because heterozygosity has low heritability and inbreeding avoidance means there is no single ‘best’ male for all females.

Females usually invest much more in offspring than males and are therefore expected to be choosy, for instance favouring partners with good genes⁴. However, despite many documented examples of female choice⁵, the link between choosiness and offspring quality is not always clear. Moreover, in polygynous mating systems the importance of female choice remains controversial because it is often difficult to distinguish between choice for a top male *per se* and, for example, choice for a prime location that top males coincidentally prefer⁶. Perhaps more importantly, there is also the theoretical problem that strong selection for traits associated with quality will rapidly remove additive genetic variability for those traits and hence the benefits of choice, a problem termed the paradox of the lek^{1–3}. Explaining why females remain choosy despite the lek paradox therefore presents a long-standing challenge to evolutionary biologists^{7,8}. Here we help to resolve the paradox in fur seals by showing that female choice operates to create offspring with high heterozygosity, a trait that improves quality, but has low heritability.

Pinnipeds are interesting models for studying female choice because, although once considered classic examples of polygyny on the basis of male dominance, genetic studies have revealed disparities between behavioural and genetic estimates of male breeding success^{9,10}. Even in fur seals, where agreement is better^{11,12}, a female may exert choice by moving away from the territory in which she spends most time. Thus, in a sympatrically breeding mixed-species colony, females become more mobile around the time of oestrus, presumably to mate with conspecific males¹¹.

To explore the basis of female choice in a strongly polygynous pinniped, we have studied an unusually amenable system based on Antarctic fur seals *Arctocephalus gazella* breeding at Bird Island, South Georgia. Unprecedented access is provided by an aerial walkway above the colony, allowing both extensive genetic sampling and individual positions to be determined daily to one-metre accuracy. Most if not all conceptions occur on land in the colony and unambiguous paternity can be assigned to a high proportion of pups using

a panel of nine highly polymorphic microsatellites¹². Remarkably, genetic recapture data reveal that whereas females are reasonably mobile, males rarely move more than a body-length (2 m) during the main part of the breeding season¹³. Anecdotally, we have observed individual females shifting their position within the colony around the time of oestrus, only to return subsequently to their pupping locations. Accordingly, our genetic data show that whereas 42% of females conceive to males within a radius of two body lengths (3 m, Fig. 1), appreciable numbers conceive to males up to 35 m away, suggesting that active choice is occurring. Moreover, only 23.5% of females ($n = 73$) conceive to the closest male.

Previously, we have shown that microsatellite heterozygosity is a significant predictor of male reproductive quality in terms of traits such as dominance, longevity and ability to hold territories in harsh years when pup production is low¹⁴. We therefore constructed a general linear model (GLM) to test the hypothesis that females that move further conceive to more heterozygous males, estimated as internal relatedness, IR¹⁵ (see Methods), with mother-to-father relatedness fitted as an additional explanatory factor. We find a highly significant negative relationship between distance moved and male IR (Table 1; Fig. 2; $P < 0.001$), indicating that females are prepared to move further to find a more heterozygous partner. This does not appear to be an artefact of colony structure, for example females moving to where low IR males congregate, because scrambling the link between male identity and genotype completely removes

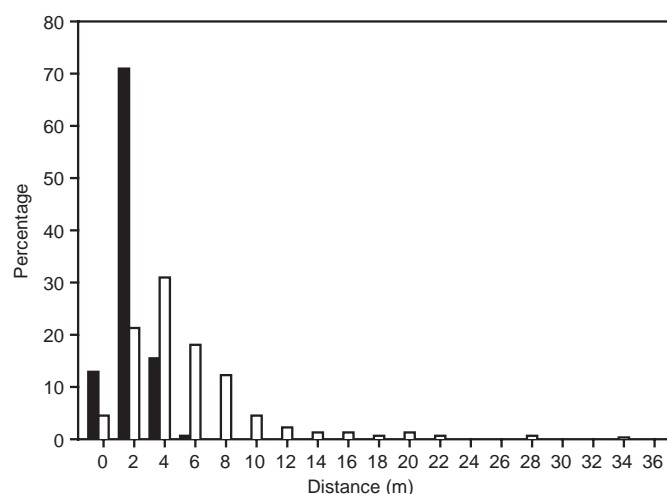


Figure 1 | Distances between male and female seals on the estimated day of conception. Black bars show the distance of each female to her nearest male, and white bars show the distance of each female to her pup's father ($n = 310$ genetically assigned paternities).

¹Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK. ²British Antarctic Survey, High Cross, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

Table 1 | General linear model of distance moved by female, fitting father's internal relatedness (IR) and mother–father relatedness as explanatory variables

Variable	Estimate	d.f.	χ^2	P
Father's IR	−7.3357	1	7.3592	0.0008
Mother–father relatedness	−0.1710	1	2.7315	0.0672
Interaction (father's IR * mother–father relatedness)	36.6956	1	5.0572	0.0255

Data were restricted to 237 females that did not mate with their nearest male (see Methods). Total deviance = 5060.2; explained deviance = 6.6%; $n = 237$. d.f., degrees of freedom.

significance. In addition, the interaction of male IR with male–female relatedness is also significant ($P = 0.026$, Table 1), suggesting that heterozygous males are less attractive if they are closely related to the female. Again, scrambling the link between female identity and genotype removes this pattern.

Heterozygosity can impact fitness either through genome-wide patterns exposing deleterious recessive alleles (inbreeding depression) or through single locus heterosis¹⁶. To test for the latter, we fitted observed heterozygosity at each of the loci (as categorical variables with heterozygotes scored as one and homozygotes scored as zero) and found that every one yields a similarly non-significant but consistently positive slope (Supplementary Table 1; Sign test, $P = 0.004$). Such a pattern, with the effect spread evenly among loci rather than concentrated in one or two, suggests that the mechanism predominantly involves genome-wide (inbreeding) effects.

In many studies of polygynous species, unambiguous demonstration of female choice is difficult, even when it seems the most likely explanation of what is happening. Our system is unusual in that males, although dispersed throughout the breeding colony, remain essentially static¹³. In contrast, females tend to be more mobile and can exert choice by visiting favoured males in a way that is perhaps analogous to a dispersed lek. Movement is difficult across the crowded colony, and potentially costly both because a female has to leave her pup unattended and owing to aggression from other animals, particularly females that object to any intrusion by an adult because crushing is a major source of pup mortality¹⁷. Consequently,

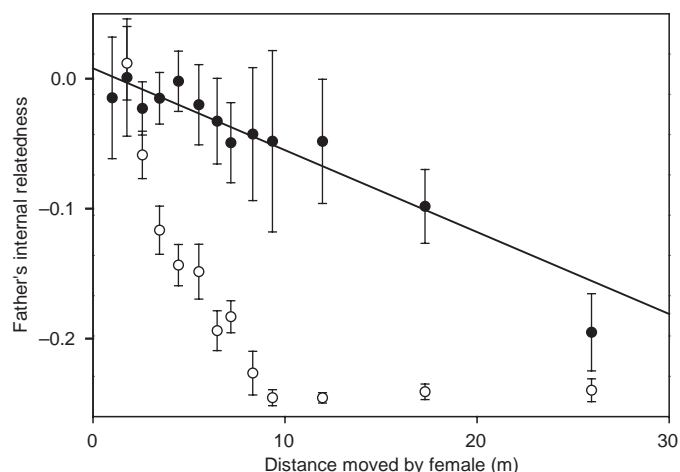


Figure 2 | Relationship between distance moved by female and father's internal relatedness (IR, filled circles). Data were restricted to 237 females that did not mate with their nearest male (see Methods). Owing to decreasing sample sizes with increasing distance moved, data are summarized (mean \pm s.e.) over 1 m intervals for the first 10 metres, then over two successive 5 m intervals and finally over the remaining 15 m. The trend line (linear regression) is given for guidance only. Open circles show the mean \pm s.e. minimum IR of males present within a radius of the distance moved. No data were available for the 1 m interval because the only males that were closer to the female than the father were unsampled.

as we observe, larger movements should be made only in return for greater genetic benefits.

We find that females choose males according to a balance between high heterozygosity and low relatedness, both of which will tend to increase offspring heterozygosity. Thus, whereas male IR predicts offspring heterozygosity (calculated as parental relatedness rather than offspring IR to maximize genetic resolution; $r^2 = 2.9\%$, $n = 310$, $P = 0.003$), female IR does not ($r^2 = 0.03\%$, $n = 310$, $P = 0.853$). In fur seals, although heterozygosity does not affect pre-weaning survival because most mortality is due to starvation and accidental crushing¹⁷, high heterozygosity has a positive influence on many aspects of male reproductive success¹⁴ and may well reduce disease and parasite burdens¹⁸, indicating that females that exert choice in this way will produce fitter offspring. At the same time, our findings may help to resolve the lek paradox in two ways. First, although a few of the most heterozygous males gain many paternities, the low heritability of heterozygosity means that variability in preferred traits will not be lost. Second, because females modulate their choice according to their relatedness to potential mates, different females often favour different males. Indeed, females that move far almost invariably move more than they need simply to maximize the heterozygosity of their partner (Fig. 2). Whether additive variability is also important is uncertain, though preference for heterozygosity would tend to reduce the strength of selection on this component and hence reduce its rate of loss.

Previous solutions to the lek paradox are somewhat unsatisfactory and largely lack direct empirical support. One class of model avoids loss of additive variability by proposing low heritability for the selected trait, through either high polygenicity or environment-dependent expression^{7,19}. However, low heritability implies proportionately lower benefits gained by choosy females. Similarly, choice for genetic compatibility²⁰ could alleviate the problem of most females choosing the same male, but has yet to be demonstrated for multilocus systems, most empirical evidence coming from the MHC^{21,22}. Finally, theory indicates that selection for heterozygous partners does not solve the paradox when based on a single locus and has not been examined for multilocus scenarios^{23,24}. In contrast, by choosing according to both multilocus heterozygosity and relatedness, female fur seals seem able to raise offspring quality in a way that should not exhaust additive variability for fitness.

How are females able to assess male genetic traits remotely? Heterozygosity could potentially be assessed visually through body size and/or condition²⁵, dominant behaviours²⁶, or territory quality²⁷. Relatedness can be assessed olfactorily using cues from the MHC²¹, but it is difficult to imagine how a female could detect genome-wide relatedness. Consequently, we refitted the GLMs, each time substituting relatedness calculated using one locus for multilocus-relatedness. Interestingly, one locus is individually significant (locus *Lw10* interaction term, $P = 0.006$), all other loci being non-significant ($P > 0.3$ in all cases). Because it is unlikely that one locus by chance lies adjacent to the MHC (if adjacent is ± 5 Mb, for the human genome $P = 0.029$), our data raise the possibility that females assess relatedness using one or more MHC-like loci. As yet, we are unsure whether females are able to detect olfactory cues across a crowded colony, or instead assess males before the breeding season and remember favoured individuals through either sight or their vocalizations. Elucidating the exact mechanisms by which this is achieved should provide a fertile area for further research.

METHODS

Study site and observational data. This study was conducted at Bird Island, South Georgia (54° 00' S, 38° 02' W) during the austral summers of 1994/1995–2003/2004. The study population was located at a small cobblestone breeding beach covering an area of 440 m² and with an average pup production of 668 during the study. A scaffold walkway provided access to all parts of the beach, enabling animals to be observed and tissue-sampled with minimal disturbance.

Approximately 700 randomly selected adult females were fitted with plastic cattle ear tags (Dalton Supplies, Henley-on-Thames). Because adult territorial

males were too large and aggressive to be captured, they were instead identified using distinctive natural markings (for example, scars) and by applying small patches of gloss paint to the pelage²⁸. Throughout each season, daily surveys were made of all pups born to tagged females on the study beach, in which birth locations were recorded to the nearest square metre relative to grid markings painted on the scaffold walkway. We also recorded the daily locations of all adult males that held territories on the study beach.

Tissue sampling and microsatellite genotyping. Piglet ear-notching pliers were used to collect a small skin sample from the interdigital margin of the foreflipper of tagged females and their pups. Adult males were remotely sampled using a biopsy dart system. Genomic DNA was extracted and genotyped using nine dinucleotide-repeat microsatellite loci as described elsewhere²⁹. The loci exhibited clear banding patterns, did not deviate significantly from Hardy–Weinberg equilibrium and were highly polymorphic, yielding up to 18 alleles per locus. The genotyping error rate, assessed by independently re-genotyping 190 individuals at all nine loci, was low at 0.0038 per reaction or 0.0022 per allele²⁹.

Parentage assignment, heterozygosity and relatedness. Parentage was assigned to 569 pups using a strict exclusion approach implemented in the program NEWPAT as described by Hoffman *et al.*¹². Importantly, paternity matches were nearly identical to those assigned using a likelihood-based approach and the exclusion probability was sufficiently high (0.999964) that nearly every assignment was likely to be correct¹². We quantified heterozygosity at nine microsatellite loci using the measure internal relatedness, IR¹⁵. IR estimates the relatedness of an individual's parents using the extent of allele sharing relative to random expectations by adapting the method of Queller and Goodnight³⁰. Although IR is strongly correlated with standardized heterozygosity¹⁴, it is theoretically more informative because it weights allele sharing by the frequencies of those alleles. In addition, we also calculated Queller and Goodnight's³⁰ relatedness between each pup's mother and father.

Statistical analyses. Spatial analyses were restricted to pups for which the mother's pupping location was known and the genetically assigned father was present on the beach within one day of the conception date ($n = 310$), estimated as the pupping day plus seven days²⁸. We used the program R (<http://www.r-project.org>) to construct a general linear model (GLM) of distance moved by females, fitting male IR and mother–father relatedness as explanatory terms, together with first order interactions. Using standard deletion testing procedures, each term was dropped from the model unless doing so significantly reduced the amount of deviance explained. The change in deviance between the full and reduced models was distributed as χ^2 with degrees of freedom equal to the difference in the degrees of freedom between the models with and without the term in question. Because single-order terms cannot be removed from models where significant higher-order interactions are present, the significance of lower-order terms was assessed by comparison with reduced models not containing higher-order interactions. Additionally, the most parsimonious model was selected as having the lowest Akaike information criterion (AIC).

Females that mate with the nearest male may do so either because they are not choosy, or because they have exercised choice when looking where to pup and have settled close to a favoured partner. If the latter occurs, females that mate with their nearest male could include both individuals exerting no choice and those expressing strong choice, potentially confounding our analysis. Consequently the models presented use a restricted data set of 237 females that did not mate with their nearest male. However, all analyses were repeated using the full data set and yield very similar results (Supplementary Table 2). If anything, models for the full data set yield lower P -values (father's IR = 0.0006, mother–father relatedness = 0.0449, interaction = 0.0144) but explain less deviance (5.1%), supporting the notion that some females may pup near preferred partners.

Received 3 November; accepted 22 December 2006.
Published online 7 February 2007.

- Borgia, G. In *Sexual selection and reproductive competition in insects* (eds Blum, M. S. & Blum, N. A.) 19–80 (Academic Press, New York, 1979).
- Taylor, P. D. & Williams, G. C. The lek paradox is not resolved. *Theor. Popul. Biol.* **22**, 392–409 (1982).
- Kirkpatrick, M. & Ryan, M. The evolution of mating preferences and the paradox of the lek. *Nature* **350**, 33–38 (1991).
- Trivers, R. L. in *Sexual selection and the descent of man, 1871–1971* (ed. Campbell, B.) 136–179 (Aldine-Atherton, Chicago, 1972).
- Andersson, M. & Iwasa, Y. Sexual selection. *Trends Ecol. Evol.* **11**, A53–A58 (1996).
- Birkhead, T. & Moller, A. Female control of paternity. *Trends Ecol. Evol.* **8**, 100–104 (1993).
- Tomkins, J. L., Radwan, J., Kotiaho, J. S. & Tregenza, T. Genic capture and resolving the lek paradox. *Trends Ecol. Evol.* **19**, 323–328 (2004).

- Qvarnström, A., Brommer, J. E. & Gustafsson, L. Testing the genetics underlying the co-evolution of mate choice and ornament in the wild. *Nature* **441**, 84–86 (2006).
- Worthington Wilmer, J., Allen, P. J., Pomeroy, P. P., Twiss, S. D. & Amos, W. Where have all the fathers gone? An extensive microsatellite analysis of paternity in the grey seal (*Halichoerus grypus*). *Mol. Ecol.* **8**, 1417–1429 (1999).
- Lidgard, D. C., Boness, D. J., Bowen, W. D., McMillan, J. I. & Fleischer, R. C. The rate of fertilization in male mating tactics of the polygynous grey seal. *Mol. Ecol.* **13**, 3543–3548 (2004).
- Goldsworthy, S. D., Boness, D. J. & Fleischer, R. C. Mate choice among sympatric fur seals: female preference for conphenotypic males. *Behav. Ecol. Sociobiol.* **45**, 253–267 (1999).
- Hoffman, J. I., Boyd, I. L. & Amos, W. Male reproductive strategy and the importance of maternal status in the Antarctic fur seal *Arctocephalus gazella*. *Evolution Int. J. Org. Evolution* **57**, 1917–1930 (2003).
- Hoffman, J. I., Trathan, P. N. & Amos, W. Genetic tracking reveals extreme site fidelity in territorial male Antarctic fur seals *Arctocephalus gazella*. *Mol. Ecol.* **15**, 3841–3847 (2006).
- Hoffman, J. I., Boyd, I. L. & Amos, W. Exploring the relationship between parental relatedness and male reproductive success in the Antarctic fur seal *Arctocephalus gazella*. *Evolution Int. J. Org. Evolution* **58**, 2087–2099 (2004).
- Amos, W. *et al.* The influence of parental relatedness on reproductive success. *Proc. R. Soc. Lond. B* **268**, 2021–2027 (2001).
- Hansson, B. & Westerberg, L. On the correlation between heterozygosity and fitness in natural populations. *Mol. Ecol.* **11**, 2467–2474 (2002).
- Hoffman, J. I., Forcada, J. & Amos, W. No relationship between microsatellite variation and neonatal fitness in Antarctic fur seals, *Arctocephalus gazella*. *Mol. Ecol.* **15**, 1995–2005 (2006).
- Acevedo-Whitehouse, K., Gulland, F., Grieg, D. & Amos, W. Inbreeding: Disease susceptibility in California sea lions. *Nature* **422**, 35 (2003).
- Rowe, L. & Houle, D. The lek paradox and the capture of genetic variance by condition dependent traits. *Proc. R. Soc. Lond. B* **263**, 1415–1421 (1996).
- Neff, B. D. & Pitcher, T. E. Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Mol. Ecol.* **14**, 19–38 (2005).
- Brown, J. L. & Eklund, A. Kin recognition and the major histocompatibility complex: an integrative review. *Am. Nat.* **143**, 435–461 (1994).
- Bernatchez, L. & Landry, C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J. Evol. Biol.* **16**, 363–377 (2003).
- Partridge, L. in *Mate Choice* (ed. Bateson, P.) 227–256 (Cambridge University Press, Cambridge, 1983).
- Irwin, A. J. & Taylor, P. D. Heterozygous advantage and the evolution of female choice. *Evol. Ecol. Res.* **2**, 119–128 (2000).
- Pujolar, J. M., Maes, G. E., Vancollie, C. & Volckaert, F. A. M. Growth rate correlates to individual heterozygosity in the European eel, *Anguilla anguilla* L. *Evolution Int. J. Org. Evolution* **59**, 189–199 (2005).
- Tiira, K. *et al.* Do dominants have higher heterozygosity? Social status and genetic variation in brown trout, *Salmo trutta*. *Behav. Ecol. Sociobiol.* **59**, 657–665 (2006).
- Seddon, N., Amos, W., Mulder, R. A. & Tobias, J. A. Male heterozygosity predicts territory size, song structure and reproductive success in a cooperatively breeding bird. *Proc. R. Soc. Lond. B* **271**, 1823–1829 (2004).
- Arnould, J. P. Y. & Duck, C. D. The cost and benefits of territorial tenure, and factors affecting mating success in male Antarctic fur seals. *J. Zool.* **241**, 649–664 (1997).
- Hoffman, J. I. & Amos, W. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Mol. Ecol.* **14**, 599–612 (2005).
- Queller, D. C. & Goodnight, K. F. Estimating relatedness using genetic markers. *Evolution* **43**, 258–275 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Briggs, S. Robinson, M. Jessop, K. Reid, R. Taylor, T. Walker and N. Warren for help with logistics and fieldwork. We are also grateful to T. Clutton-Brock, S. Hodge and K. Isvaran for helpful comments on the manuscript. This work was funded by the Natural Environment Research Council (NERC) and the Antarctic Funding Initiative (AFI), and contributes to the BAS DISCOVERY 2010 science programme. J.I.H. was also funded by the Isaac Newton Trust, Balfour fund and Cambridge University VIP scheme. Fieldwork was approved by BAS and the University of Cambridge Animal Ethics Board. Samples were collected and retained under permits issued by the Department for Environment, Food and Rural Affairs (DEFRA), and in accordance with the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES).

Author Contributions J.I.H. contributed to the concept, genotyping, data analysis, and writing; W.A. to the concept, data analysis and writing; and J.F. and P.N.T. to field project coordination and the collection of tissue samples and observational data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.I.H. (jjh24@cam.ac.uk).

An African origin for the intimate association between humans and *Helicobacter pylori*

Bodo Linz^{1*}, François Balloux^{2*}, Yoshan Moodley¹, Andrea Manica³, Hua Liu², Philippe Roumagnac¹, Daniel Falush⁴, Christiana Stamer¹, Franck Prugnolle⁵, Schalk W. van der Merwe⁶, Yoshio Yamaoka⁷, David Y. Graham⁷, Emilio Perez-Trallero⁸, Torkel Wadstrom⁹, Sebastian Suerbaum¹⁰ & Mark Achtman¹

Infection of the stomach by *Helicobacter pylori* is ubiquitous among humans. However, although *H. pylori* strains from different geographic areas are associated with clear phylogeographic differentiation^{1–4}, the age of an association between these bacteria with humans remains highly controversial^{5,6}. Here we show, using sequences from a large data set of bacterial strains that, as in humans, genetic diversity in *H. pylori* decreases with geographic distance from east Africa, the cradle of modern humans. We also observe similar clines of genetic isolation by distance (IBD) for both *H. pylori* and its human host at a worldwide scale. Like humans, simulations indicate that *H. pylori* seems to have spread from east Africa around 58,000 yr ago. Even at more restricted geographic scales, where IBD tends to become blurred, principal component clines in *H. pylori* from Europe strongly resemble the classical clines for Europeans described by Cavalli-Sforza and colleagues⁷. Taken together, our results establish that anatomically modern humans were already infected by *H. pylori* before their migrations from Africa and demonstrate that *H. pylori* has remained intimately associated with their human host populations ever since.

Over half of all humans are infected by *Helicobacter pylori*, a Gram-negative bacterium that can cause peptic ulcers and constitutes a risk factor for stomach cancer⁸. Not only is *H. pylori* ubiquitous, but it also possesses strong phylogeographic structure¹, suggesting that bacterial polymorphisms reflect human phylogeography and historical migrations^{2,3,5}. In 2003, we assigned 370 *H. pylori* strains to four main population clusters, two of which were subdivided into subpopulations². The geographic sources of these strains reflected major events in human settlement history, such as the colonisation of Polynesia and the Americas and the African Bantu migrations. However, these discrete groupings seem to contradict an apparent continuity of the geographic component of genetic diversity in humans: genetic differentiation between human populations increases linearly with geographic distance computed along landmasses^{9–12}; and their genetic diversity declines with increasing geographic distance from east Africa^{13,14}.

There are several possible explanations why detailed population genetic patterns differ between *H. pylori* and their human hosts. Infection of humans by *H. pylori* might be too recent to have been affected by ancient events in human settlement history, for example, it might date from a recently acquired zoonosis⁵. Or differences in

population structure between bacteria and humans may reflect frequent horizontal transmission of *H. pylori*. Alternatively, apparent differences in the population genetic patterns may simply be a matter of perception owing to differing analytical methodology: *H. pylori* population genetics has so far focused on the description of clusters, whereas human population genetics is influenced by a traditional emphasis on clines¹⁵.

We used an expanded data set (769 *H. pylori* isolates from 51 ethnic sources; Supplementary Table 1) to test whether patterns in their geographic distribution mimic those of humans. Bayesian MCMC (Monte-Carlo Markov chain) cluster analyses identified the same five ancestral sources of nucleotides as found previously with a smaller data set² (Fig. 1a, b). These analyses also assigned the isolates to six populations containing various degrees of ancestry from the five ancestral sources (Fig. 1a; Supplementary Fig. 1a). Four of the populations had been previously identified, and designated hpEurope, hpEastAsia, hpAfrica1 and hpAfrica2 owing to their obvious geographical associations². In agreement, almost all strains isolated from Europeans belong to hpEurope, including Basques in Spain, Russians and Kazakhs, and most isolates from east Asia belong to hpEastAsia (Supplementary Table 1). The results also confirmed that hpAfrica2, previously represented by only few isolates, is common in South Africa. Two new populations were identified and designated hpAsia2 and hpNEAfrica. hpAsia2 was isolated in northern India, Thailand, Bangladesh, the Philippines and elsewhere in southeastern Asia (Supplementary Fig. 1b). hpNEAfrica was predominant among isolates from Ethiopia, Somalia, Sudan and Nilo-Saharan speakers in northern Nigeria.

Matrices of pairwise F_{ST} (a measure of genetic differentiation between populations) between paired groups of samples from analogous geographic locations were strongly correlated (Mantel regression coefficient = 0.86, $P < 0.001$) between *H. pylori* sequences and human microsatellite data¹⁶; 73% of human variation can be explained by a linear relationship with microbial variation (Supplementary Fig. 2). Thus, the geographic component of genetic diversity seems to be quantitatively comparable between *H. pylori* and humans, except that F_{ST} is considerably higher in *H. pylori*.

We next address evidence for a continuum of genetic ancestry in *H. pylori*. Whereas the assignments of individual isolates to populations are quite unambiguous with the no-admixture model in STRUCTURE¹⁷, its linkage model shows that the proportions of ancestry from the five

¹Department of Molecular Biology, Max-Planck Institut für Infektionsbiologie, D-10117 Berlin, Germany. ²Theoretical and Molecular Population Genetics Group, Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK. ³Evolutionary Ecology Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. ⁴Department of Statistics, University of Oxford, Oxford OX1 3SY, UK. ⁵Génétique et Evolution des Maladies Infectieuses, UMR IRD-CNRS 2724, centre IRD de Montpellier, 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 05, France. ⁶Department of Internal Medicine and Gastroenterology, University of Pretoria, Pretoria 0002, South Africa. ⁷Department of Medicine–Gastroenterology, Baylor College of Medicine and Michael E. DeBakey VA Medical Center, Houston, TX 77030, USA. ⁸Department of Microbiology, Donostia Hospital, 20014 San Sebastian, Spain. ⁹Department of Laboratory Medicine, Lund University, SE22632 Lund, Sweden. ¹⁰Medizinische Hochschule Hannover, Institut für Medizinische Mikrobiologie und Krankenhaushygiene, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany.

*These authors contributed equally to this work.

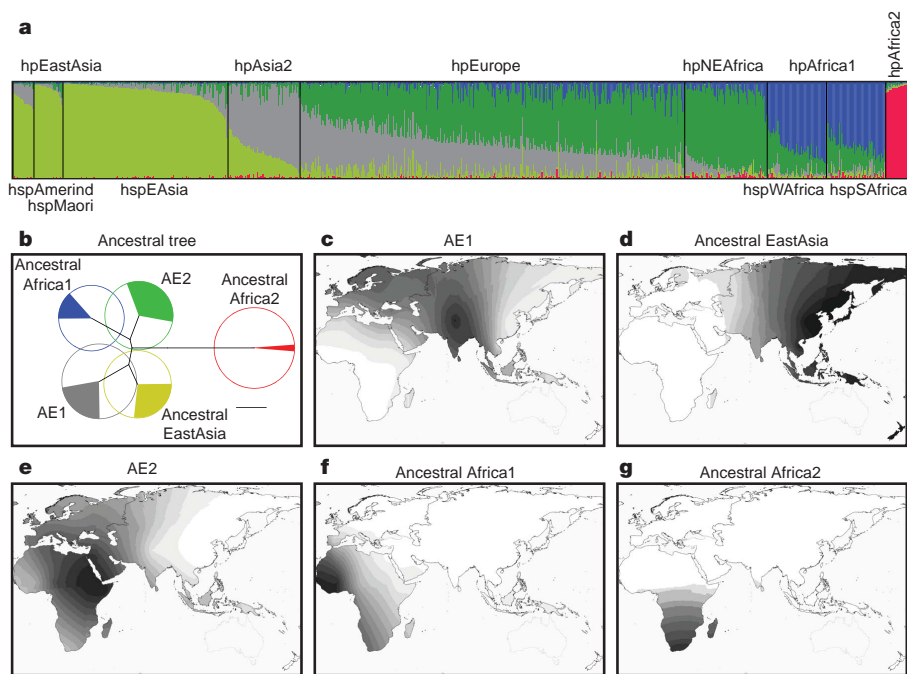


Figure 1 | Five ancestral populations in *H. pylori*. **a**, DISTRUCT²⁴ plot of the proportions of ancestral nucleotides in 769 *H. pylori* isolates as determined by Structure V2.0 (linkage model)¹⁷. A thin line for each isolate indicates the estimated amount of ancestry from each ancestral population as five coloured segments (see **b** for colour code). The lines are grouped by (sub)population and ordered by ancestry. **b**, Neighbour-joining tree (black lines) of the relationships between and diversity within five ancestral populations calculated as described². Circle diameters are proportional to

within-population genetic diversity and angles of filled arcs are proportional to the amount of ancestry attributable to each population among modern strains. Scale bar, 0.01. **c–g**, Spatial distribution of five geographic sources of ancestral nucleotides. Dark–light gradients show clinal declines in proportions of ancestral nucleotides by distance from a geographic centre. Colour-coding as in **b**. AE1, AE2, ancestral Europe 1 and 2, respectively. hspAmerind, hspMaori, hspEAsia are subpopulations of hpEastAsia; hspSAfrica and hspWAFrica are subpopulations of hpAfrica1.

ancestral sources almost form a continuum (Fig. 1a) between the five populations other than hpAfrica2, which is highly distinct. Similarly, although most concatenated sequences cluster according to their population assignment in a phylogenetic tree (Supplementary Fig. 3), their relatedness is also almost continuous, again with the exception of hpAfrica2. A nearly continuous distribution of the proportions of ancestry suggests localized admixture due to recombination. Such admixture would blur differences between initially distinct populations in close geographical proximity, and could potentially lead to strong signals of IBD in *H. pylori*, as observed in humans^{9–12}.

Diversity was only poorly correlated with geographical sources in initial analyses, which might reflect noise due to recent human migrations plus horizontal transmission of *H. pylori* between ethnically distinct groups. Therefore, we excluded 147 isolates from obvious recent human migrants and their admixed descendants as well as 31 isolates whose population assignments were highly incongruent with their sources of isolation (horizontal transmission) (see Methods). The resulting ‘non-migrant’ data set (Supplementary Table 2) contained 532 *H. pylori* isolates and 1,405 polymorphisms. The results were compared with human diversity based on 783 autosomal microsatellites¹⁶. Similar to previous analyses^{11,12,14}, 77% of the variance in F_{ST} between autosomal human markers from distinct geographic sources was accounted for by the shortest geographic distance along landmasses (Fig. 2a). For *H. pylori*, only 47% of the variance was accounted for by geographic distance (Fig. 2b, $P \leq 0.001$), but this estimate rose to 72% when a standard conversion of genetic diversity was plotted against the logarithm of the geographic distance for the 442 haplotypes from geographic locations with at least 10 isolates (Supplementary Fig. 4). Thus, comparable proportions of the genetic diversity are due to IBD in *H. pylori* as in humans.

Genetic diversity within modern humans decreases with distance from east Africa, reflecting their recent African origin^{12–14}; 85% of this decrease in diversity could be accounted for by distance from east

Africa (Fig. 2c). The non-migrant *H. pylori* data set also showed a similar trend (Fig. 2d, $P \leq 0.001$) and 59% of the decrease in diversity could be accounted for by distance from east Africa. Unlike IBD, where trends with *H. pylori* might mimic those of humans without a joint demography, parallel decreases in diversity with distance from east Africa indicate close associations between the two. Simulations

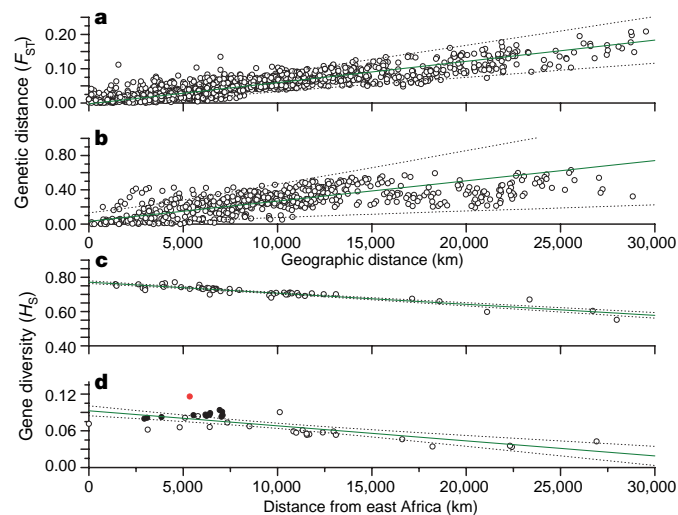


Figure 2 | Parallel geographic patterns of genetic diversity in humans and *H. pylori*. **a, b**, Genetic distance in humans (**a**) and *H. pylori* (**b**) between pairs of geographic populations (F_{ST}) versus geographic distance between the two populations. **c, d**, Average gene diversity in humans (**c**) and *H. pylori* (**d**) within geographic populations (H_S) versus geographic distance from east Africa. R^2 is 0.77 for **a**; 0.47 for **b**; 0.85 for **c**; and 0.59 for **d**. In **d**, samples that are predominantly composed of hpEurope isolates are indicated by filled circles, whereas the red circle identifies the sample from South Africa. Confidence intervals are indicated by dotted lines.

Table 1 | Simulations for three hypothetical scenarios for the origin of *H. pylori*.

Scenario	Estimated parameters					Model	
	Ancestral population (K_0)	Carrying capacity (K)	Migration (K^*m)	Growth rate (r)	Time (years)	R^2	AIC
East Africa	561 \pm 182	203 \pm 87	65 \pm 6	0.73 \pm 0.14	57,955 \pm 3,748	0.68	83.3
South Africa	541 \pm 214	185 \pm 76	86 \pm 8	0.68 \pm 0.13	61,746 \pm 4,436	0.57	88.1
East Asia	747 \pm 207	467 \pm 191	370 \pm 169	0.70 \pm 0.16	36,500 \pm 6,728	0.02	101.5

Additional details are in Supplementary Information. The five estimated parameters are the size of the founder population (K_0) and the carrying capacity of any subsequently colonized demes (K), both of which are expressed as the number of effective strains that succeed in infecting the subsequent host generation. K^*m represents the number of effective migrants sent by any deme, r is the growth rate within demes in a logistic growth model, and time refers to the duration of the entire colonization process. The time estimates presented here include an additional 2,500 yr migration phase after the last deme had been colonized. Mean and standard deviations were computed for all simulations that fell within the 95% confidence interval. The fit of the models to the data are given by R^2 , the amount of variance explained, and AIC, the Akaike information criterion.

with human data indicated that they migrated from east Africa 56,000 \pm 5,500 yr ago¹⁴. Similar demographically explicit genetic simulations now indicate that an east African source for *H. pylori* is more probable than South Africa or China, and that *H. pylori* migrated from east Africa 58,000 \pm 3,500 yr ago (Table 1, Supplementary Table 3 and Supplementary Fig. 5; see Supplementary Information for details). Thus, we conclude that *H. pylori* accompanied anatomically modern humans during their migrations from Africa that have been estimated at 50,000–70,000 yr ago^{14,18–21}. This implies that *H. pylori* was present in Africa before these migrations, suggesting that Africa is the source of both *H. pylori* and humans.

Are clines in genetic diversity truly contradictory to discrete clusters? Discrete clusters can be defined even within a perfectly continuous pattern owing to sampling artefacts¹⁰. But for *H. pylori*, geographic isolation is only a marginally better predictor of genetic differentiation than discrete clusters based on genetic similarity. Within a generalized linear model framework, assignment of populations to the six clusters defined by STRUCTURE explains 70% of the variance of pairwise F_{ST} for populations with 10 or more isolates, versus up to 72% by geographical distance. This situation resembles previous results for the geographic apportionment of human genetic diversity^{11,16}, except that even when geography is first accounted for, clusters as defined by STRUCTURE still explain 11% of additional variance in genetic differentiation between *H. pylori* populations as compared with only 2% in humans. We therefore examined the modern geographic sources of the nucleotides associated with the five ancestral populations according to STRUCTURE (Fig. 1c–g). The spatial distribution of ancestral nucleotides indicated that ancestral Europe2 (AE2) originated in east Africa, AE1 in central Asia, ancestral EastAsia in east Asia and ancestral Africa1 and Africa2 in Africa. These data probably reflect extensive population expansions, subsequent to the global spread that accompanied migrations out of Africa, and may well reflect important episodes in human history during the Neolithic period and later.

If *H. pylori* were also a marker for human migrations at a more local scale, one would expect to find similar patterns between human and bacterial diversity within Europe, which was not one of the sources of ancestral nucleotides in *H. pylori*. Indeed, the first two principal components of spatial autocorrelation with hpEurope isolates (Fig. 3a, b) were very similar to those that had been obtained with human allozymes in classical work by Cavalli-Sforza and colleagues⁷ (Fig. 3d, e) and the third principal component (PC) showed similar east–west clines. Such clines were originally interpreted as genetic signatures or indications of episodic migratory events⁷, although this interpretation has been questioned²². We note that the first principal component of the *H. pylori* data, PC1, is a cline from the southeast that correlates ($R^2 = 0.35$, $P < 0.01$) with the proportion of ancestry from AE2 (Supplementary Fig. 6a), which originated in northeastern Africa. For PC2, a cline from the northeast, the correlation ($R^2 = 0.6$, $P < 0.01$) is with AE1 (Supplementary Fig. 6b), which originated in Central Asia. These correlations show that in *H. pylori*, as previously suggested², much of the spatial pattern observed in Europe can be attributed to admixture from different sources. It also supports the controversial hypothesis⁷ that similar clines in humans are also due to waves of migration of genetically

distinct populations into Europe (demic diffusion), except that the spatial sources of ancestral nucleotides are assigned to northeastern Africa and Central Asia. We further conclude that there are highly striking, quantitative parallels in clines and IBD both at the global and the local scale between humans and *H. pylori*. These presumably reflect the dissemination of *H. pylori* by a variety of prehistoric human migrations, followed by admixture after horizontal transmission between human populations.

In this paper we have shown that the key patterns in the distribution of *H. pylori* genetic diversity mirror those of its human host. At a worldwide scale, we recovered similar patterns of isolation by distance, though absolute genetic differentiation is higher in *H. pylori*. As in humans, we observed a continuous loss of genetic diversity with increasing geographic distance from east Africa, the likely cradle of anatomically modern humans. Even at the more restricted scale of Europe, we largely recreated the classical clines described by Cavalli-Sforza and colleagues. Finally, simulations predict that *H. pylori* has spread from east Africa over the same time scale as anatomically modern humans. These extraordinary parallel population genetic patterns between *H. pylori* and its human host all demonstrate an old association predating the ‘out of Africa’ event⁴. The results further point to a scenario where *H. pylori* and human

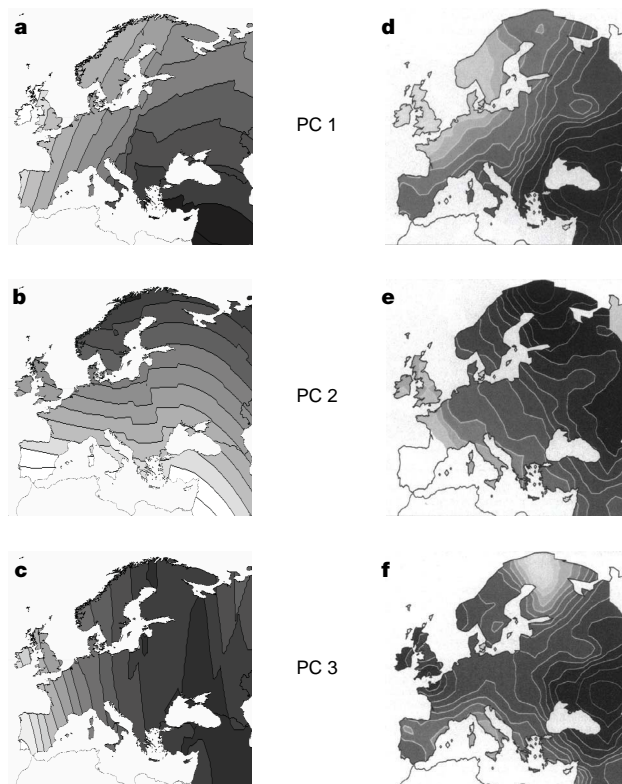


Figure 3 | Similar clinal gradients between principal components 1–3 from European *H. pylori* and humans. a–c, *H. pylori* concatenated sequences. d–f, Human allozymes. Panels d–f, reprinted with permission from citation 7 (copyright 1995 National Academy of Sciences, USA).

populations have been evolving intimately ever since, with limited long-range transmission by horizontal infections.

METHODS

Bacterial isolates and sequencing. The expanded *H. pylori* data set consists of 3,406 base pairs (bp) of unique, concatenated sequences of fragments of *atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI* and *ypcC* from 769 *H. pylori* isolates (Supplementary Table 1). The data set includes 347 novel isolates in addition to data from 422 other strains that have been described previously^{2,3,23}. The new bacteria were isolated from 25 additional ethnic sources in Asia (8 countries), Europe (4 countries, including Basques), Africa and the Middle East (9 countries) and South America (2 countries), for a total of 51 ethnic sources (Supplementary Table 1). The forward and reverse strands were sequenced as described¹. Almost half (1,522 sites, 45%) of the nucleotides are polymorphic, resulting in a nucleotide diversity (π) of 4.2% for the entire data set.

The non-migrant data set excluded bacteria that were isolated from the following migrant human populations: Europeans and Cape Coloureds from Cape Town; Mestizos from Colombia and Venezuela; Whites and African Americans from the USA; isolates in Thailand from Chinese or without ethnic association. hpAfrica2 isolates from Xhosa near Pretoria were excluded because they were a selective subset rather than a population-wide sample. The Philippines' samples were also removed because almost all bacterial populations were found there, probably owing to their colonial history. For isolates from Native Americans, only hspAmerind strains were considered non-migrant. The data set was further restricted to geographic samples with at least four isolates, to avoid statistical noise, which resulted in the elimination of all Jewish and Russian isolates and singletons from locations in China and Japan.

Other methods and details are to be found in Supplementary Information.

Received 4 October; accepted 22 December 2006.

Published online 7 February 2007.

1. Achtman, M. *et al.* Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol. Microbiol.* **32**, 459–470 (1999).
2. Falush, D. *et al.* Traces of human migrations in *Helicobacter pylori* populations. *Science* **299**, 1582–1585 (2003).
3. Wirth, T. *et al.* Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proc. Natl Acad. Sci. USA* **101**, 4746–4751 (2004).
4. Eppinger, M. *et al.* Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genet.* **2**, e120 (2006).
5. Kersulyte, D. *et al.* Differences in genotypes of *Helicobacter pylori* from different human populations. *J. Bacteriol.* **182**, 3210–3218 (2000).
6. Dailidiene, D. *et al.* *Helicobacter acinonychis*: genetic and rodent infection studies of a *Helicobacter pylori*-like gastric pathogen of cheetahs and other big cats. *J. Bacteriol.* **186**, 356–365 (2004).
7. Piazza, A. *et al.* Genetics and the origin of European languages. *Proc. Natl Acad. Sci. USA* **92**, 5836–5840 (1995).
8. Suerbaum, S. & Michetti, P. *Helicobacter pylori* infection. *N. Engl. J. Med.* **347**, 1175–1186 (2002).

9. Relethford, J. H. Global patterns of isolation by distance based on genetic and morphological data. *Hum. Biol.* **76**, 499–513 (2004).
10. Serre, D. & Paabo, S. Evidence for gradients of human genetic diversity within and among continents. *Genome Res.* **14**, 1679–1685 (2004).
11. Manica, A., Prugnolle, F. & Balloux, F. Geography is a better determinant of human genetic differentiation than ethnicity. *Hum. Genet.* **118**, 366–371 (2005).
12. Ramachandran, S. *et al.* Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl Acad. Sci. USA* **102**, 15942–15947 (2005).
13. Prugnolle, F., Manica, A. & Balloux, F. Geography predicts neutral genetic diversity of human populations. *Curr. Biol.* **15**, R159–R160 (2005).
14. Liu, H., Prugnolle, F., Manica, A. & Balloux, F. A geographically explicit genetic model of worldwide human-settlement history. *Am. J. Hum. Genet.* **79**, 230–237 (2006).
15. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, Princeton, New Jersey, 1994).
16. Rosenberg, N. A. *et al.* Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet.* **1**, e70 (2005).
17. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
18. Underhill, P. A. *et al.* Y chromosome sequence variation and the history of human populations. *Nature Genet.* **26**, 358–361 (2000).
19. Ingman, M. & Gyllenstein, U. Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *J. Hered.* **92**, 454–461 (2001).
20. Zhivotovsky, L. A., Rosenberg, N. A. & Feldman, M. W. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am. J. Hum. Genet.* **72**, 1171–1186 (2003).
21. Mellars, P. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc. Natl Acad. Sci. USA* **103**, 9381–9386 (2006).
22. Currat, M. & Excoffier, L. The effect of the Neolithic expansion on European molecular diversity. *Proc. Biol. Sci.* **272**, 679–688 (2005).
23. Momynaliev, K. T. *et al.* Population identification of *Helicobacter pylori* isolates from Russia. *Genetika* **41**, 1182–1185 (2005).
24. Rosenberg, N. A. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**, 137–138 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the receipt of bacterial strains from A. van der Ende, M. J. Blaser, N. J. Saunders, R. J. Owen, F. Mégraud and sequences from K. T. Momynaliev and C. Kraft. We thank J. Goudet for providing a modified version of FSTAT able to deal with the large data set and help with R by K. -P. Pleissner. Grant support was from the German Federal Ministry for Education and Research (BMBF) in the framework of the PathoGenoMik Network (M.A., S.S.), the Biotechnology and Biological Sciences Research Council (F.B.), the Swedish Research council (T.W.) and Lund University Hospital (T.W.).

Author Information EMBL accession numbers for DNA sequences, AM413111–418360. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A. (achtman@mpiib-berlin.mpg.de), F.B. (fb255@mole.bio.cam.ac.uk) or S.S. (Suerbaum.Sebastian@mh-hannover.de).

Planning for the future by western scrub-jays

C. R. Raby¹, D. M. Alexis¹, A. Dickinson¹ & N. S. Clayton¹

Knowledge of and planning for the future is a complex skill that is considered by many to be uniquely human. We are not born with it; children develop a sense of the future at around the age of two and some planning ability by only the age of four to five^{1–3}. According to the Bischof-Köhler hypothesis⁴, only humans can dissociate themselves from their current motivation and take action for future needs: other animals are incapable of anticipating future needs, and any future-oriented behaviours they exhibit are either fixed action patterns or cued by their current motivational state. The experiments described here test whether a member of the corvid family, the western scrub-jay (*Aphelocoma californica*), plans for the future. We show that the jays make provision for a future need, both by preferentially caching food in a place in which they have learned that they will be hungry the following morning and by differentially storing a particular food in a place in which that type of food will not be available the next morning. Previous studies have shown that, in accord with the Bischof-Köhler hypothesis, rats⁵ and pigeons⁶ may solve tasks by encoding the future but only over very short time scales. Although some primates and corvids^{7–9} take actions now that are based on their future consequences, these have not been shown to be selected with reference to future motivational states¹⁰, or without extensive reinforcement of the anticipatory act¹¹. The results described here suggest that the jays can spontaneously plan for tomorrow without reference to their current motivational state, thereby challenging the idea that this is a uniquely human ability.

Evidence from both brain-damaged and healthy humans suggests that two forms of mental time travel, retrospective in the case of episodic memory and prospective in the case of future planning, depend on common neuropsychological processes^{1,12–14}. On the basis of the observation that western scrub-jays store and recover food caches in the wild, experimental studies have shown that their ability to recover their caches depends on an episodic-like memory for the caching episode. Specifically, jays remember what food they have cached, where and when it was cached¹⁵, and which other birds observed their caching¹⁶. These memories are then used flexibly, both to guide their recovery of the food caches and to protect their food caches against being stolen by other birds¹⁷. To the extent that episodic memory and future planning depend on common processes, the caching behaviour of these birds should reflect an ability to anticipate future need states. To assess this prediction, we investigated whether scrub-jays can plan for a future motivational need, as opposed to a current one. To do so, we gave eight birds experience of two different compartments on alternate mornings for six days. In one compartment they were always given breakfast and in the other they were not. After this training, the jays were unexpectedly given food to eat and cache in the evening. If they were capable of forward thinking, they should have cached more food in the compartment in which they had not been given breakfast and therefore would expect to be hungry the next morning, relative to the compartment in which they had been given breakfast.

Each bird was housed in three adjoining compartments, A, B and C (Fig. 1). These compartments could be divided from each other or left open so that a bird had access to all three spaces. In the 'planning for breakfast' experiment, on training days, having not eaten during the night, each bird was shut in compartment A or C on alternate mornings for two hours with identical substrate-filled caching trays. In one compartment (the 'breakfast' compartment) they received a breakfast of powdered pine nuts, which they could eat but not cache, and in the other they were not given breakfast (the 'no-breakfast' compartment). Throughout the rest of the day they had access to all compartments with food freely available. Two hours before darkness, the caching trays were again placed in compartments A and C and the birds were deprived of food for 90 min. In compartment B, they were then given powdered, non-cacheable pine nuts, which they could eat freely for the 30 min before darkness. During this period the birds continued to have access to all three compartments. After each bird had experienced three 'no-breakfast' training trials and three 'breakfast' training trials, they were tested one evening for their ability to anticipate the future by replacing the powdered food with whole, cacheable pine nuts, thereby giving them for the first time the opportunity to cache in the trays in compartments A and C for the morning, as well as eating the food immediately.

The birds anticipated their hunger the next morning by storing significantly more pine nuts in the caching tray in the 'no-breakfast' compartment (16.3 ± 1.8 ; mean \pm s.e.m.) than in the 'breakfast' compartment (5.4 ± 1.8 ; mean \pm s.e.m.) (paired *t*-test with 7 degrees of freedom, $t_7 = 3.01$, $P = 0.02$).

We considered the possibility that the differential caching was due to a propensity to cache in places associated with hunger. Rats have been shown to eat more in a room previously associated with hunger¹⁸, suggesting that hunger can be conditioned to a particular context. Consequently, a second experiment was devised to contrast a conditioned hunger account with an explanation in terms of future planning. In the 'breakfast choice' experiment, the birds were always given breakfast in the morning but the food that they received differed depending on their location; for example, the jays were always given dog kibble on mornings when they were confined to compartment A (the 'kibble-for-breakfast' compartment) and peanuts on

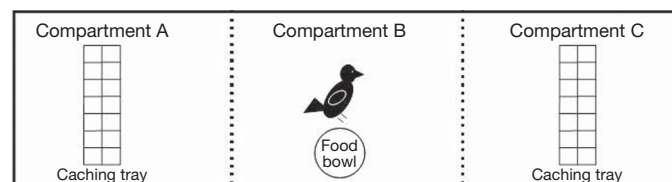


Figure 1 | Compartmental layout used for the 'planning for breakfast' experiment. The position of caching trays is shown in compartments A and C, and of the food bowl in compartment B. Dotted lines represent the compartmental divisions, although during caching no dividers were in place. In the second experiment, the compartmental layout was the same except that two food bowls, equidistant from compartments A and C, were used.

¹Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.

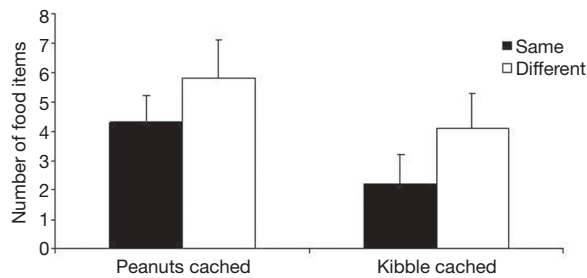


Figure 2 | Mean number of peanuts and kibble cached in the 'breakfast choice' experiment. The scenario in which peanuts were cached in the 'peanuts-for-breakfast' compartment and kibble was cached in the 'kibble-for-breakfast' compartment is denoted as 'same'. The scenario in which peanuts were cached in the 'kibble-for-breakfast' compartment and kibble was cached in the 'peanuts-for-breakfast' compartment is denoted as 'different'. The jays cached significantly more items of the food type that was different to the food that they had previously received for breakfast in that compartment relative to the number of items they cached of the food that was the same as they received for breakfast in that compartment ($F_{(1,8)} = 5.48$; $P = 0.047$). There was no overall difference between the amount of food cached in each compartment ($F < 1$), nor did the jays cache either food type more than the other overall ($F_{(1,8)} = 2.29$, non-significant), ($n = 9$). Error bars, \pm s.e.m.

mornings that they were confined to compartment C (the 'peanuts-for-breakfast' compartment), or vice versa. Apart from this difference, the experimental procedure was identical to the first experiment except that in the evening both foods were offered in powdered form in compartment B. It has been shown that associating a place with a particular food potentiates consumption of that specific food in that location¹⁹ and that caching and eating by scrub-jays are controlled by a common motivational system²⁰. Therefore a conditioning account would predict that the jays would cache a particular food in the compartment previously associated with that food. If, however, jays prefer a choice of food at breakfast and are capable of forward planning, they should have preferentially cached the 'different' food rather than the 'same' food in each compartment when offered both foods in cacheable form.

In accord with the future planning account, and contrary to the conditioning account, at test the birds stored significantly more of the 'different' food than the 'same' food in each compartment relative to the amount of that food that they stored in the other compartment. The results and statistics are summarized in Fig. 2.

These results challenge the Bischof-Köhler hypothesis⁴ by demonstrating that caching on one day was controlled by the next day's motivational state and available resources. The birds were mildly hungry when given food to cache as this promotes caching, but they ate food at the same time as caching. If they were caching for current hunger, they would have no reason to cache preferentially in one compartment rather than the other, or to cache one food rather than the other in a particular compartment. In the absence of language, there is no knowing whether this reflects episodic future thinking, in which the bird is projecting itself into tomorrow morning's situation, or semantic future thinking, in which the jay takes prospective action, but without personal mental time travel into the future. However, in either case it shows that these birds must have the capability to plan for a future motivational state over a timescale stretching at least into tomorrow. These results, therefore, challenge the assumption that the ability to anticipate and take action for future needs evolved only in the hominid lineage.

METHODS

Subjects and apparatus in the 'planning for breakfast' experiment. Five female and three male adult western scrub-jays were used, all of whom had participated in a suite of experiments on episodic-like memory¹⁷ and prospective cognition⁹. They were housed in pairs in a row of four linked compartments, each measuring 2 m \times 1 m \times 1 m. One of each pair of birds was tested each week, and training

and testing were conducted in three of the compartments (A, B and C in Fig. 1). During an experiment the test subject was periodically confined to one of the compartments, which was divided from the other two by clear plastic dividers so the other two compartments of the housing remained visible. The bird that was not participating was located in the fourth adjoining compartment, separated from the other bird by opaque dividers.

The jays were maintained under artificial light on a 12:12 h light-dark cycle and fed a maintenance diet of shelled peanuts, sunflower seeds, grains, and dog biscuits. During experimental training and tests they were given pine nuts presented in 14.5-cm-diameter plastic bowls. During training the pine nuts were in a non-cacheable powdered form (mixed with some powdered maintenance diet to avoid 'clumping') to ensure that the birds only cached during the experimental test. Birds cached in plastic ice-cube trays (11 cm wide \times 26 cm long), which consisted of an array of 16 cube moulds (4.5 cm \times 3 cm \times 3 cm) filled with corn cob, each of which was a potential cache site. Each tray was attached to a wooden board (15 cm wide \times 30 cm long).

Procedure for the 'planning for breakfast' experiment. Each bird received six training trials followed by one test trial. From 09:00–17:00 each day the birds' maintenance diet was freely available. At 17:00 the test subject was separated from its companion into compartments A, B and C. Maintenance diet was removed for 90 min at 17:00 and caching trays were put in compartments A and C. At 18:30 a bowl containing 10 g of powdered pine nuts, which could be eaten but not cached, was put in the centre of compartment B for 30 min. At 19:00 the bowl of pine nuts and the caching trays were removed. The birds were deprived of food overnight so that they were mildly hungry in the morning. Each morning at 07:00 the experimental bird was confined in either compartment A or C for two hours. The caching trays were returned to both compartments A and C for this period. In one compartment the bird was given no food ('no-breakfast' compartment) and in the other the bird was given powdered pine nuts ('breakfast' compartment). Each bird experienced the two compartments on alternate days, with the compartment that was the 'breakfast' compartment (A or C) counterbalanced across birds. The order in which each bird experienced the two compartments was also counterbalanced. The test day followed exactly the same evening routine as training days with the exception that 30 (cacheable) whole pine nuts were given at 18:30 instead of powdered food. The jays did not get the opportunity to recover their caches on the morning after the test trial.

Subjects and apparatus in the 'breakfast choice' experiment. Two adult jays, one male and one female, were added to those tested in the first experiment. However, one was excluded for failing to approach either caching tray on test, leaving a total of nine birds. The birds' maintenance diet was changed to exclude peanuts and dog kibble, and to include Harrison's high potency fine organic bird food. The experimental set-up was identical to that described for the first experiment in all but one respect. The jays were fed peanuts and dog kibble in two separate bowls during the training and testing, instead of pine nuts.

Procedure for the 'breakfast choice' experiment. The experimental protocol was exactly the same as for the first experiment in all but four respects. On training days at 18:30 two bowls, one containing 10 g of powdered peanuts and one containing 10 g of powdered dog kibble were put in the centre of compartment B, equidistant from compartments A and C for 30 min. In the mornings in one of compartments A or C the birds were given powdered peanuts ('peanut' compartment) and in the other they were given powdered dog kibble ('kibble' compartment). Birds were counterbalanced both with respect to the order in which they experienced each food and with respect to each compartment. Tests followed exactly the same evening routine as on training days with the exception that 15 whole dog kibble and 15 whole peanuts were given at 18:30 instead of powdered food.

Analysis. The caching data in the first experiment were analysed using a paired *t*-test and in the second experiment using a repeated measures ANOVA. Alpha was set at 0.05. Parametric tests were used because our data conform to the standard assumptions of homogeneity of variance and normality. In the 'planning for breakfast' experiment, $n = 8$, in the 'breakfast choice' experiment, $n = 9$.

Received 22 November 2006; accepted 8 January 2007.

- Atance, C. M. & O'Neill, D. K. Episodic future thinking. *Trends Cogn. Sci.* 5, 533–539 (2001).
- Atance, C. M. & Meltzoff, A. N. My future self: young children's ability to anticipate and explain future states. *Cogn. Dev.* 20, 341–361 (2005).
- Suddendorf, T. & Busby, J. Making decisions with the future in mind: developmental and comparative identification of mental time travel. *Learn. Motiv.* 36, 110–125 (2005).
- Suddendorf, T. & Corballis, M. C. Mental time travel and the evolution of the human mind. *Genet. Soc. Gen. Psychol. Monogr.* 123, 133–167 (1997).
- Cook, R. C., Brown, M. F. & Riley, D. A. Flexible memory processing by rats: use of prospective and retrospective information in the radial maze. *J. Exp. Psychol. Anim. Behav. Process.* 11, 453–469 (1985).

6. Zentall, T. R., Steirn, J. N. & Jackson-Smith, P. Memory strategies in pigeons' performance of a radial-arm-maze analog task. *J. Exp. Psychol. Anim. Behav. Process.* **16**, 358–371 (1990).
7. Mulcahy, N. J. & Call, J. Apes save tools for future use. *Science* **312**, 1038–1040 (2006).
8. Roberts, W. A. Are animals stuck in time? *Psychol. Bull.* **128**, 473–489 (2002).
9. Clayton, N. S., Dally, J., Gilbert, J. & Dickinson, A. Food caching by western scrub-jays (*Aphelocoma californica*) is sensitive to the conditions at recovery. *J. Exp. Psychol. Anim. Behav. Process.* **31**, 115–124 (2005).
10. Suddendorf, T. Foresight and evolution of the human mind. *Science* **312**, 1006–1007 (2006).
11. Naqshbandi, M. & Roberts, W. A. Anticipation of future events in squirrel monkeys (*Saimiri sciureus*) and rats (*Rattus norvegicus*): tests of the Bischof-Kohler hypothesis. *J. Comp. Psych.* **120**, 345–357 (2006).
12. Rosenbaum, R. S. *et al.* The case of KC: contributions of a memory-impaired person to memory theory. *Neuropsychologia* **43**, 989–1021 (2005).
13. Klein, S. B., Loftus, J. & Kihlstrom, J. F. Memory and temporal experience: the effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Soc. Cogn.* **20**, 353–379 (2002).
14. Okuda, J. *et al.* Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *Neuroimage* **19**, 1369–1380 (2003).
15. Clayton, N. S. & Dickinson, A. Episodic-like memory during cache recovery by scrub jays. *Nature* **395**, 272–274 (1998).
16. Dally, J. M., Emery, N. J. & Clayton, N. S. Food-caching western scrub-jays keep track of who was watching when. *Science* **312**, 1662–1665 (2006).
17. Clayton, N. S., Bussey, T. J. & Dickinson, A. Can animals recall the past and plan for the future? *Nature Rev. Neurosci.* **4**, 685–691 (2003).
18. Roitman, M. F., van Dijk, G., Thiele, T. E. & Bernstein, I. L. Dopamine mediation of the feeding response to violations of spatial and temporal expectancies. *Behav. Brain Res.* **122**, 193–199 (2001).
19. Petrovich, G. D., Ross, C. A., Gallagher, M. & Holland, P. C. Learned contextual cue potentiates eating in rats. *Physiol. Behav.* (in the press)
20. Clayton, N. S. & Dickinson, A. Motivational control of caching behaviour in the scrub jay, *Aphelocoma coerulescens*. *Anim. Behav.* **57**, 435–444 (1999).

Acknowledgements This research was supported by a BBSRC Grant and the University of Cambridge, and was conducted within an MRC Cooperative Grant. Thanks to J. Dally, A. Seed and M. Ellis for comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.S.C. (nsc22@cam.ac.uk).

LETTERS

eIF4E function in somatic cells modulates ageing in *Caenorhabditis elegans*

Popi Syntichaki¹, Kostoula Troulinaki¹ & Nektarios Tavernarakis¹

Regulation of protein synthesis is critical for cell growth and maintenance. Ageing in many organisms, including humans, is accompanied by marked alterations in both general and specific protein synthesis¹. Whether these alterations are simply a corollary of the ageing process or have a causative role in senescent decline remains unclear. An array of protein factors facilitates the tight control of messenger RNA translation initiation². The eukaryotic initiation factor 4E (eIF4E), which binds the 7-monomethyl guanosine cap at the 5' end of all nuclear mRNAs, is a principal regulator of protein synthesis³. Here we show that loss of a specific eIF4E isoform (IFE-2) that functions in somatic tissues⁴ reduces global protein synthesis, protects from oxidative stress and extends lifespan in *Caenorhabditis elegans*. Lifespan extension is independent of the forkhead transcription factor DAF-16, which mediates the effects of the insulin-like signalling pathway on ageing. Furthermore, IFE-2 deficiency further extends the lifespan of long-lived *age* and *daf* nematode mutants. Similarly, lack of IFE-2 enhances the long-lived phenotype of *clk* and dietary-restricted *eat* mutant animals. Knockdown of target of rapamycin (TOR), a phosphatidylinositol kinase-related kinase that controls protein synthesis in response to nutrient cues, further increases the longevity of *ife-2* mutants. Thus, signalling via eIF4E in the soma is a newly discovered pathway influencing ageing in *C. elegans*.

The molecular mechanisms underlying age-associated changes in protein synthesis are largely unknown. In eukaryotes, the rate of cap-dependent protein synthesis is mainly determined by the translation initiation factor eIF4E³. We examined the role of eIF4E in *C. elegans* ageing. The *C. elegans* genome encodes five eIF4E isoforms (IFE-1 to IFE-5), which differ in cap-binding specificity and anatomical expression. Biochemical studies demonstrate that IFE-2 is the only eIF4E isoform in the soma that binds both 7-monomethyl guanosine and 2,2,7-trimethyl guanosine caps, present on nematode mRNAs. IFE-4 and the germline-specific IFE-3 only bind 7-monomethyl guanosine caps. IFE-1 and IFE-5, which are present in the germ line, also bind both 7-monomethyl guanosine and 2,2,7-trimethyl guanosine caps⁴.

RNA interference (RNAi) knockdown of *ife-2* resulted in substantial extension of nematode lifespan (Fig. 1a; see also Supplementary Table 1). We observed similar lifespan extension in animals carrying an *ife-2* deletion (Fig. 1a). Depletion of other eIF4E isoforms had no effect on longevity (Fig. 1b; see also Supplementary Table 1). *ife-2* encodes the most abundantly expressed eIF4E isoform in *C. elegans* somatic tissues (Fig. 1c; see also Supplementary Fig. 1 and ref. 5). No *ife-2* expression is detected in the germ line, where *ife-1*, *ife-3* and *ife-5* are expressed^{4,5}. *ife-4* encodes an additional eIF4E isoform expressed in the soma⁴. Consistent with RNAi experiments, we did not detect lifespan extension in animals carrying an *ife-4* deletion (Fig. 1b). We also measured the longevity of animals deficient for both somatic-specific eIF4E isoforms. These nematodes showed lifespan equal to *ife-2(RNAi)* knockdown animals (Supplementary Table 1).

To establish further the somatic origin of longevity conferred by IFE-2 depletion, we used animals carrying a temperature-sensitive mutation in *glp-4* (*abnormal germline proliferation-4*), a zygotic gene required for normal post-embryonic proliferation of the germ line⁶. These mutants are germline-deficient at the restrictive temperature (25 °C). Similar to wild-type animals, knockdown of *ife-2* extends the lifespan of *glp-4* mutants (Fig. 1d; see also Supplementary Table 1). Therefore, lack of germ line does not suppress the effect of IFE-2 depletion on animal lifespan. We conclude that elimination of a specific eIF4E isoform, expressed in somatic cells, extends lifespan in *C. elegans*.

We considered whether IFE-2 depletion results in any other obvious defects that might account for the effects on animal lifespan. We examined both *ife-2* knockout mutants and wild-type animals subjected to *ife-2* RNAi for feeding behaviours, fertility, developmental abnormalities, movement defects and anatomical alterations. IFE-2-depleted animals grown at 20 °C were indistinguishable from wild-type animals for pharyngeal pumping of bacterial food into the intestine, defecation rhythms, fecundity, developmental timing, sinusoidal locomotion, body size and shape, and internal organ morphology (gonad, intestine, musculature and pharynx). IFE-2 deficiency induced significant embryonic lethality at 25 °C, whereas fecundity was relatively unaffected (Supplementary Table 2). We also tested the capacity of IFE-2-depleted animals to become dauer larvae, a stress-resistant, developmentally arrested larval form, induced by adverse environmental conditions. We did not observe any effects on dauer formation as a result of IFE-2 deficiency (Supplementary Table 2). In addition, we find that *ife-2* mutants are indistinguishable from wild-type animals for heat-shock resistance (Supplementary Fig. 2).

Ageing in *C. elegans* is mainly regulated by a conserved endocrine signalling pathway that involves the insulin/insulin-like growth factor (IGF) receptor DAF-2 (*abnormal dauer formation-2*) and the phosphatidylinositol-3-OH kinase (PI(3)K) AGE-1 (*ageing alteration-1*). Mutations that compromise the activity of DAF-2 or AGE-1 extend animal lifespan. Longevity conferred by *daf-2* and *age-1* mutations requires the forkhead box, sub-group O (FOXO) transcription factor DAF-16 (reviewed in ref. 7). We find that IFE-2 depletion further extends the lifespan of long-lived *daf-2* and *age-1* mutants (Fig. 2a; see also Supplementary Table 1). DAF-16 is not required for lifespan extension by IFE-2 deficiency (Fig. 2b; see also Supplementary Table 1). This observation indicates that IFE-2 functions downstream or independently of DAF-16 to control ageing. To distinguish between these two possibilities, we tested whether *ife-2* transcription is under the control of DAF-16. Expression of a full-length IFE-2::GFP fusion, driven by the *ife-2* promoter, is not affected by *daf-16* knockdown (Fig. 2c). Furthermore, depletion of DAF-2, which modulates DAF-16 activity via a protein phosphorylation cascade, does not affect *ife-2* expression (Supplementary Fig. 4). Therefore, *ife-2* is not a downstream target of DAF-16.

¹Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, Heraklion 71110, Crete, Greece.

Mutations in *clk-1* (*clock abnormality-1*), a gene required for the biosynthesis of ubiquinone (an essential component of the mitochondrial electron transport chain), also extend *C. elegans* lifespan⁸. Knockdown of *ife-2* further extends the lifespan of long-lived *clk-1*

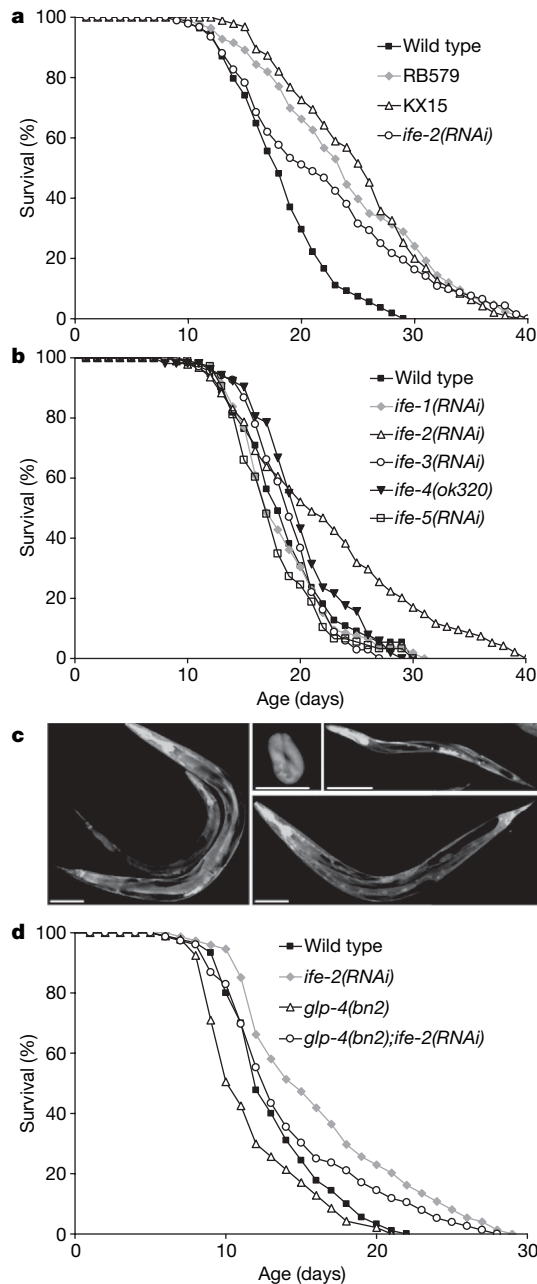


Figure 1 | eIF4E deficiency in somatic tissues extends *C. elegans* lifespan. The percentage of animals remaining alive is plotted against animal age. Lifespan values are given in Supplementary Table 1. **a**, Depletion of IFE-2 by gene deletion or RNAi extends animal lifespan. Strains RB579 and KX15 both harbour the *ife-2(ok306)* deletion allele. KX15 is an outcrossed version of RB579. Assays were carried out at 20 °C. **b**, Knockdown of the germline-specific *ife-1*, *ife-3* and *ife-5* genes does not alter animal lifespan. In addition, animals carrying the *ife-4(ok320)* deletion allele show normal lifespan. Assays were carried out at 20 °C. **c**, Images of animals expressing the *pife-2::IFE-2::GFP* reporter fusion. *ife-2* is highly expressed in all somatic tissues of adult animals, including the pharynx, the intestine, body wall muscles, the hypodermis, neurons and the canal cell. Expression is high during all post-embryonic developmental stages, throughout adulthood, and is detectable in twofold stage embryos (details are shown in Supplementary Fig. 1; scale bars, 100 μ m). **d**, Knockdown of *ife-2* extends the lifespan of *glp-4(bn2)* mutant animals lacking a germ line. Assays were carried out at 25 °C.

mutant animals (Fig. 2d; see also Supplementary Table 1). We observed a similar additive effect on lifespan in long-lived, dietary restricted *eat-2* (*eating, abnormal pharyngeal pumping-2*) mutants⁹ (Fig. 2e; see also Supplementary Table 1). Taken together, our results suggest that regulation of eIF4E activity in somatic tissues constitutes a novel mechanism that influences longevity in *C. elegans*.

eIF4E activity is regulated by association with inhibitory eIF4E-binding proteins (4E-BPs) and by direct phosphorylation mediated by the mitogen-activated protein (MAP) kinase signal-integrating

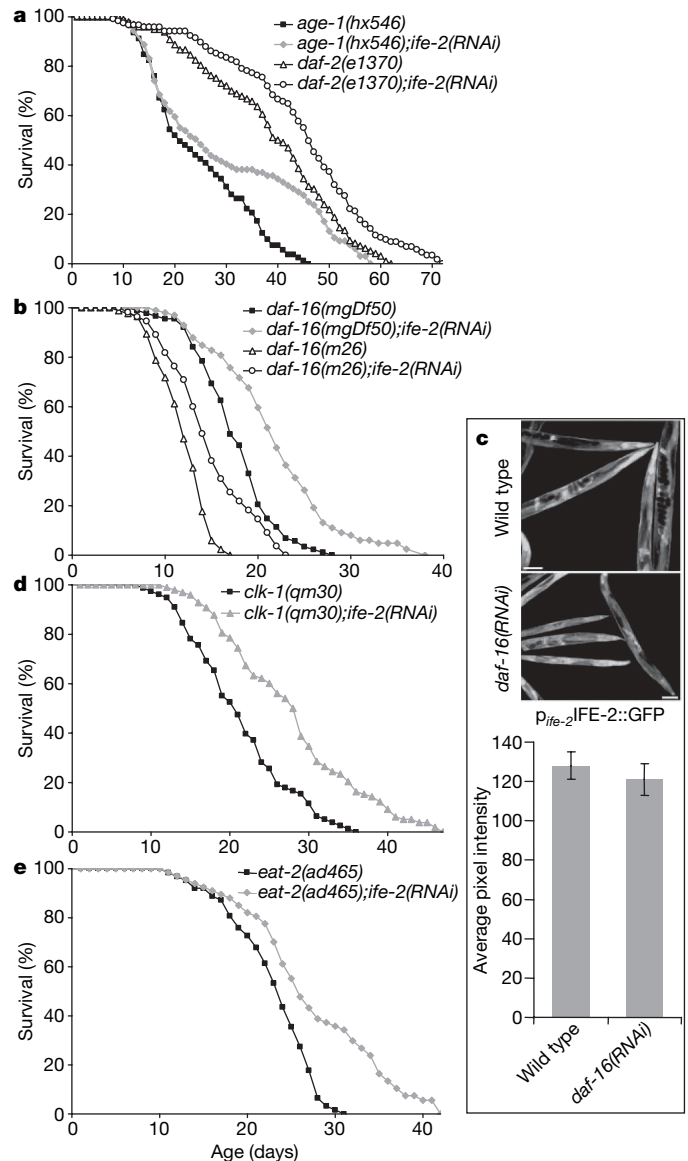


Figure 2 | IFE-2 and genes influencing ageing in *C. elegans*. Survival curves of mutant nematode populations subjected to *ife-2* RNAi are shown, with the percentage of nematodes remaining alive plotted against age. Lifespan values are given in Supplementary Table 1; assays were carried out at 20 °C. **a**, Knockdown of *ife-2* further increases the lifespan of long-lived *age-1* and *daf-2* mutants. **b**, Lifespan extension by IFE-2 depletion is not suppressed by two *daf-16* mutant alleles. **c**, Expression of *ife-2* is independent of the DAF-16 transcription factor. Knockdown of *daf-16* does not affect expression of full-length IFE-2::GFP driven by the *ife-2* promoter (*pife-2::IFE-2::GFP*). Representative images of nematodes bearing the *pife-2::IFE-2::GFP* transgene subjected to *daf-16* RNAi compared to wild-type controls are shown. Quantification of animal fluorescence is also shown (five independent experiments, 50 animals total for each strain; error bars denote standard deviation; $P > 0.5$, unpaired *t*-test). **d**, **e**, Knockdown of *ife-2* further increases the lifespan of long-lived *clk-1* mutants (**d**) and long-lived *eat-2* mutant animals (**e**).

kinases (Mnk1/2)¹⁰. Phosphorylation of 4E-BP by the nutrient-sensing kinase TOR reduces its affinity for eIF4E and is thought to promote mRNA translation (reviewed in ref. 2). Notably, TOR deficiency, which dampens the rate of translation in yeast and mammals, extends *C. elegans* lifespan¹¹. This observation is consistent with our findings that eIF4E downregulation confers increased longevity. Are the effects of TOR depletion on lifespan mediated by concomitant downregulation of eIF4E via 4E-BP? This scenario predicts that elimination of 4E-BP would cancel the effect that TOR deficiency has on ageing. However, no structural 4E-BP homologue is apparent in the *C. elegans* genome (P.S. and N.T., unpublished observations). Thus, we investigated the potential link between TOR signalling and eIF4E by assaying the lifespan of animals deficient for both TOR and IFE-2. Knockdown of *ife-2* further extends the lifespan of long-lived TOR-deficient mutants (Fig. 3a; see also Supplementary Table 1). Similarly, knockdown of TOR in animals carrying an *ife-2* deletion results in additional lifespan extension (Fig. 3b; see also Supplementary Table 1). The synergy between TOR and IFE-2 indicates that their effects on ageing are mediated by distinct mechanisms.

In addition to phosphorylating 4E-BP, TOR regulates S6K, a kinase that in turn phosphorylates the S6 ribosomal protein¹². We asked whether S6K mediates the effects of TOR on ageing. Knockdown of the *C. elegans* p70S6K homologue *rsk-1* and the closely related p90S6K homologue *rskn-1* has no effect on animal lifespan (Fig. 3c; see also Supplementary Table 1). Furthermore, depletion of *rsk-1* or *rskn-1* does not suppress the lifespan extension brought about by IFE-2 deficiency (Fig. 3d; see also Supplementary Table

1). These observations suggest that the effects of IFE-2 on lifespan do not require S6K.

In mammals, the Mnk1/2 kinases modulate the activity of eIF4E by phosphorylation at Ser 209 (ref. 10). In *Drosophila*, the Mnk homologue Lk6 regulates growth in response to nutrients via eIF4E^{13,14}. We examined whether Mnk activity influences lifespan through eIF4E in *C. elegans*. RNAi knockdown of the nematode Mnk homologue *mnk-1* (MAP kinase integrating kinase-1) shortened both wild-type and *ife-2* mutant lifespan (Fig. 3e; see also Supplementary Table 1). We did not detect any concomitant developmental, behavioural or anatomical defects in *mnk-1* knockdown animals. Of note, no effect on lifespan is observed as a result of *mnk-1* knockdown in *glp-4* mutant animals lacking a germ line (Fig. 3f; Supplementary Table 1), indicating that MNK-1 function in the germ line is required to maintain normal lifespan.

Recent studies implicate eIF4E regulators such as 4E-BP, TOR and Mnk in the control of development, growth and ageing¹⁵. These regulators are coupled to the insulin/IGF signalling pathway. For example, transcription of 4E-BP is under the control of FOXO in *Drosophila*, whereas in mammals, the 4E-BP PHAS-I (phosphorylated heat- and acid-stable protein I) is regulated by insulin signalling (reviewed in ref. 2). In addition, genetic studies in *C. elegans* suggest that TOR may function downstream or independently of DAF-16 to mediate the effects of DAF-2 signalling on ageing¹¹. However, the role of eIF4E in ageing remained unknown. We have shown that eIF4E deficiency in the soma, but not in the germ line, extends *C. elegans* lifespan. Our genetic analysis suggests that eIF4E could influence ageing independently of known mechanisms that involve insulin/IGF signalling, dietary restriction and respiratory chain components. Whether these mechanisms may alter protein synthesis by targeting other mRNA translation factors remains to be elucidated.

What is the mechanism underlying lifespan extension resulting from depletion of a specific eIF4E isoform? IFE-2 is the most abundant eIF4E isoform in all *C. elegans* somatic tissues and the only one in the soma that binds both 7-monomethyl guanosine and 2,2,7-trimethyl guanosine mRNA caps (Fig. 1c; see also Supplementary Fig. 1 and refs 4 and 5). Thus, we proposed that IFE-2 deficiency impairs mRNA translation initiation, resulting in protein synthesis reduction. To test this hypothesis, we developed a novel method to monitor and compare protein synthesis rates in *C. elegans*, based on fluorescence recovery after photo-bleaching (FRAP; see Supplementary Methods). We find that protein synthesis rates are lower in animals lacking IFE-2 (Fig. 4a). We used the protein synthesis inhibitor cycloheximide to confirm that the observed effects are due to alterations in protein synthesis (Fig. 4a).

Global mRNA translation is reduced in response to most, if not all, types of cellular stress. This results in notable conservation of cellular energy, given that the process of translation consumes up to an estimated 50% of the total energy, depending on the organism¹⁶. Similarly, reduction of protein synthesis by IFE-2 depletion may prolong lifespan by lowering energy demands and the associated generation of toxic metabolic by-products such as reactive oxygen species, which contribute significantly to the ageing process^{17,18}. In addition, the concomitant increase in energy availability may allow diversion of critical resources to cellular maintenance and repair processes, thus promoting organism longevity. To test this hypothesis, we challenged *ife-2(ok306)* mutants with paraquat (*N,N'*-dimethyl-4,4'-bipyridinium dichloride) and sodium azide (NaN_3). The herbicide paraquat is a generator of superoxide anions¹⁹. NaN_3 is a potent and specific inhibitor of cytochrome *c* oxidase, which is part of the mitochondrial electron transport chain complex IV²⁰. Paraquat treatment and inhibition of cytochrome *c* oxidase induce oxidative stress^{21,22}. We find that long-lived *ife-2(ok306)* mutant animals are considerably more resistant to both paraquat and NaN_3 compared with wild-type animals (Fig. 4b). We also examined the effects of IFE-2 deficiency on the survival of *mev-1(kn1)* (*abnormal methyl viologen sensitivity-1*) mutants, which lack succinate dehydrogenase

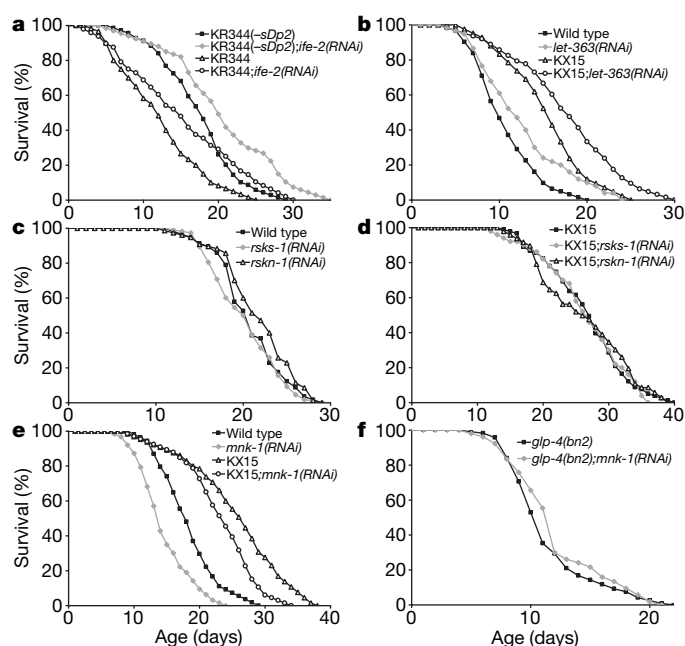


Figure 3 | Effects of regulatory kinase and IFE-2 removal on ageing. The percentage of animals remaining alive is plotted against animal age. Lifespan values are given in Supplementary Table 1. **a, b**, Knockdown of *ife-2* extends the lifespan of TOR-kinase-depleted mutants. **a**, Strain KR344 carries a loss-of-function allele of the *let-363* gene, which encodes the *C. elegans* TOR kinase, together with the chromosomal duplication *sDp2* that balances the *let-363* mutation. Loss of the *sDp2* genetic balancer ($-sDp2$) results in arrest at the L3 stage of development. Arrested *let-363* mutant animals are long lived¹¹. Isogenic animals that have not lost the duplication (KR344) are included as a control (see also Methods and Supplementary Table 1; assays were carried out at 20 °C). **b**, RNAi knockdown of *let-363* was carried out in young adult animals exiting the L4 stage. Assays were carried out at 25 °C. **c, d**, S6K kinase knockdown does not affect wild-type animal lifespan and does not suppress *ife-2* mutant longevity. Assays were carried out at 20 °C. **e**, MNK-1 deficiency shortens the lifespan of wild-type and *ife-2* mutant animals. Assays were carried out at 20 °C. **f**, *mnk-1* knockdown does not alter ageing of *glp-4* germline-deficient animals. Assays were carried out at 25 °C.

cytochrome *b*, a component of complex II of the mitochondrial electron transport chain²³. These animals are short lived and hypersensitive to oxidative stress induced by paraquat. Knockdown of *ife-2* renders *mev-1(kn1)* mutants markedly resistant to paraquat and also extends their lifespan (Fig. 4b, c). Hence, repression of protein synthesis in the soma renders animals capable of withstanding acute and chronic oxidative stress.

We next asked whether diminished IFE-2 activity in specific somatic tissues is sufficient to extend animal lifespan. We find that knockdown of *ife-2* by RNAi in all somatic cells except neurons is required to extend animal lifespan (Supplementary Fig. 3; RNAi is ineffective in the nervous system; see ref. 24). Thus, reduction of protein synthesis in neurons probably has negligible consequences on the lifespan of whole organisms.

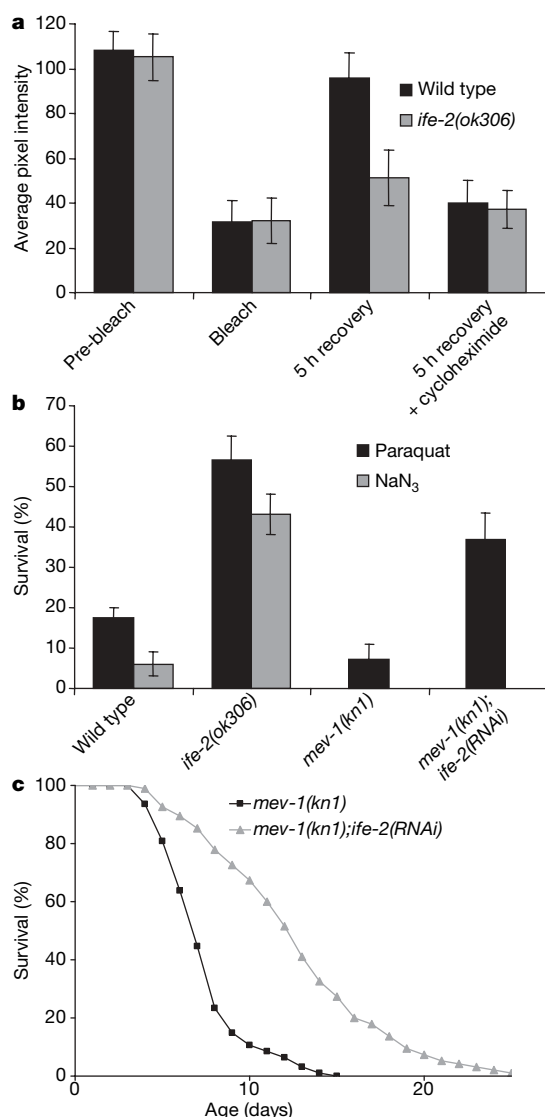


Figure 4 | eIF4E deficiency in the soma reduces protein synthesis and increases oxidative stress resistance in *C. elegans*. **a**, Quantification of fluorescence recovery after photo-bleaching in wild-type and IFE-2-deficient animals expressing GFP driven by the *ife-2* promoter (*p_{ife-2}GFP*; three independent experiments, 30 animals total for each strain; error bars denote standard deviation; $P < 0.005$, unpaired *t*-test; assays were carried out at 25 °C). **b**, Survival of synchronized animal populations under oxidative stress induced by treatment with paraquat and sodium azide (six independent experiments, 240 animals total for each strain; error bars denote standard deviation; $P < 0.005$, unpaired *t*-test). **c**, Knockdown of *ife-2* extends the lifespan of short-lived *mev-1* mutants. Assays were carried out at 20 °C.

Why is eIF4E activity in somatic tissues important for normal ageing? Studies in diverse species associate ageing with decreased protein synthesis^{25,26}. This link suggests that a sustained high rate of mRNA translation may retard senescent decline and ageing²⁷. Our work reveals a new link between protein synthesis and ageing. We demonstrate that depletion of eIF4E—a principal mRNA translation regulator—specifically in somatic cells increases oxidative stress resistance and extends lifespan. Our findings corroborate studies in mammals showing that overexpression of eIF4E promotes cellular senescence²⁸. Therefore, downregulation of mRNA translation in the soma, under appropriate conditions, may facilitate cellular maintenance and repair by moderating the large energy requirement of protein synthesis. Whereas the germ line is an immortal cell lineage²⁹, somata, which encapsulate immortal germ lines, are mortal and frail. The ‘disposable soma’ theory of ageing provides an evolutionary framework for this fundamental distinction³⁰. Failure to divert energy towards repairing stochastic damage that accumulates in the soma during life leads to inexorable decline of somatic functions and senescence. Our findings suggest that eIF4E is part of a novel mechanism, independent of insulin/IGF signalling, that modulates ageing of the soma by integrating environmental, reproductive and other cues to regulate protein synthesis and somatic maintenance.

Note added in proof: While this paper was under review, related work was published online^{31,32}.

METHODS

See Supplementary Information for detailed Methods.

Lifespan assays. Synchronous populations of nematodes were established by allowing 20 adult hermaphrodites to lay eggs for a limited time interval (4–5 h) on nematode growth medium (NGM) plates seeded with *Escherichia coli* OP50. For RNAi lifespan experiments, nematodes were placed on NGM plates containing 0.5–1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and seeded with HT115(DE3) bacteria transformed with either the pL4440 vector or the test RNAi construct. Progeny were grown on HT115-seeded plates at 20 °C unless otherwise noted, through the L4 larval stage, and then transferred to fresh HT115-seeded plates at groups of 10–20 nematodes per plate for a total of 100–150 individuals per experiment. For strains bearing the *glp-4(bn2)* mutation, lifespan assays were initiated with freshly laid eggs transferred to 25 °C, to guard against germline proliferation. The first day of adulthood was used as $t = 0$. Animals were transferred to fresh plates every 2–4 days thereafter and were examined every day for touch-provoked movement and pharyngeal pumping, until death. Each survival assay was repeated at least three times and figures represent typical experiments.

Received 5 December 2006; accepted 17 January 2007.

Published online 4 February 2007.

- Ward, W. F. The relentless effects of the aging process on protein turnover. *Biogerontology* **1**, 195–199 (2000).
- Gebauer, F. & Hentze, M. W. Molecular mechanisms of translational control. *Nature Rev. Mol. Cell Biol.* **5**, 827–835 (2004).
- Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68**, 913–963 (1999).
- Keiper, B. D. et al. Functional characterization of five eIF4E isoforms in *Caenorhabditis elegans*. *J. Biol. Chem.* **275**, 10590–10596 (2000).
- Amiri, A. et al. An isoform of eIF4E is a component of germ granules and is required for spermatogenesis in *C. elegans*. *Development* **128**, 3899–3912 (2001).
- Beanan, M. J. & Strome, S. Characterization of a germ-line proliferation mutation in *C. elegans*. *Development* **116**, 755–766 (1992).
- Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120**, 449–460 (2005).
- Lakowski, B. & Hekimi, S. Determination of life-span in *Caenorhabditis elegans* by four clock genes. *Science* **272**, 1010–1013 (1996).
- Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
- Waskiewicz, A. J. et al. Phosphorylation of the cap-binding protein eukaryotic translation initiation factor 4E by protein kinase Mnk1 in vivo. *Mol. Cell. Biol.* **19**, 1871–1880 (1999).
- Vellai, T. et al. Genetics: influence of TOR kinase on lifespan in *C. elegans*. *Nature* **426**, 620 (2003).
- Ruvinsky, I. & Meyuhas, O. Ribosomal protein S6 phosphorylation: from protein synthesis to cell size. *Trends Biochem. Sci.* **31**, 342–348 (2006).

13. Arquier, N., Bourouis, M., Colombani, J. & Leopold, P. *Drosophila* Lk6 kinase controls phosphorylation of eukaryotic translation initiation factor 4E and promotes normal growth and development. *Curr. Biol.* **15**, 19–23 (2005).
14. Reiling, J. H., Doepfner, K. T., Hafen, E. & Stocker, H. Diet-dependent effects of the *Drosophila* Mnk1/Mnk2 homolog Lk6 on growth via eIF4E. *Curr. Biol.* **15**, 24–30 (2005).
15. Jia, K., Chen, D. & Riddle, D. L. The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. *Development* **131**, 3897–3906 (2004).
16. Proud, C. G. Regulation of mammalian translation factors by nutrients. *Eur. J. Biochem.* **269**, 5338–5349 (2002).
17. Martin, G. M., Austad, S. N. & Johnson, T. E. Genetic analysis of ageing: role of oxidative damage and environmental stresses. *Nature Genet.* **13**, 25–34 (1996).
18. Stadtman, E. R. Protein oxidation and aging. *Science* **257**, 1220–1224 (1992).
19. Lin, S. J. & Culotta, V. C. The ATX1 gene of *Saccharomyces cerevisiae* encodes a small metal homeostasis factor that protects cells against reactive oxygen toxicity. *Proc. Natl Acad. Sci. USA* **92**, 3784–3788 (1995).
20. Yoshikawa, S. *et al.* Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase. *Science* **280**, 1723–1729 (1998).
21. Vanfleteren, J. R. Oxidative stress and ageing in *Caenorhabditis elegans*. *Biochem. J.* **292**, 605–608 (1993).
22. Sohal, R. S. & Weindruch, R. Oxidative stress, caloric restriction, and aging. *Science* **273**, 59–63 (1996).
23. Ishii, N. *et al.* A mutation in succinate dehydrogenase cytochrome b causes oxidative stress and ageing in nematodes. *Nature* **394**, 694–697 (1998).
24. Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A. & Driscoll, M. Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nature Genet.* **24**, 180–183 (2000).
25. Makrides, S. C. Protein synthesis and degradation during aging and senescence. *Biol. Rev. Camb. Philos. Soc.* **58**, 343–422 (1983).
26. Rattan, S. I. Synthesis, modifications, and turnover of proteins during aging. *Exp. Gerontol.* **31**, 33–47 (1996).
27. Tavernarakis, N. & Driscoll, M. Caloric restriction and lifespan: a role for protein turnover? *Mech. Ageing Dev.* **123**, 215–229 (2002).
28. Ruggero, D. *et al.* The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis. *Nature Med.* **10**, 484–486 (2004).
29. Ahmed, S. & Hodgkin, J. MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. *Nature* **403**, 159–164 (2000).
30. Kirkwood, T. B. & Austad, S. N. Why do we age? *Nature* **408**, 233–238 (2000).
31. Pan, K. Z. *et al.* Inhibition of mRNA translation extends lifespan in *C. elegans*. *Ageing Cell* doi:10.1111/j.1474-9726.2006.00266.x (5 December 2006).
32. Hansen, M. *et al.* Lifespan extension by conditions that inhibit translation in *C. elegans*. *Ageing Cell* doi:10.1111/j.1474-9726.2006.00267.x (5 December 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. Pasparaki, H. Kontaki, N. Kourtis and M. Papadakis for help with experiments. We thank M. Artal-Sanz and G. Thireos for comments on the manuscript. Some nematode strains used in this work were provided by the *C. elegans* Gene Knockout Project at OMRF (<http://www.mutantfactory.ouhsc.edu/>), which is part of the International *C. elegans* Gene Knockout Consortium, and the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR). We thank A. Fire for plasmid vectors. This work was funded by grants from EMBO and the EU sixth Framework Programme to N.T. N.T. is an EMBO Young Investigator.

Author Contributions P.S., K.T. and N.T. performed experiments; N.T. designed experiments, analysed data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.T. (tavernarakis@imbb.forth.gr).

APOBEC3 inhibits mouse mammary tumour virus replication *in vivo*

Chioma M. Okeoma¹, Nika Lovsin², B. Matija Peterlin³ & Susan R. Ross¹

Genomes of all mammals encode *apobec3* genes, which are thought to have a function in intrinsic cellular immunity to several viruses including human immunodeficiency virus type 1 (HIV-1)¹. APOBEC3 (A3) proteins are packaged into virions and inhibit retroviral replication in newly infected cells, at least in part by deaminating cytidines on the negative strand DNA intermediates². However, the role of A3 in innate resistance to mouse retroviruses is not understood. Here we show that A3 functions during retroviral infection *in vivo* and provides partial protection to mice against infection with mouse mammary tumour virus (MMTV). Both mouse A3 and human A3G proteins interacted with the MMTV nucleocapsid in an RNA-dependent fashion and were packaged into virions. In addition, mouse A3-containing and human A3G-containing virions showed a marked decrease in titre. Last, A3^{-/-} mice were more susceptible to MMTV infection, because virus spread was more rapid and extensive than in their wild-type littermates.

Mammalian genomes encode several *apobec* genes. For example, humans express A1, A2, AID and A3A–A3H proteins. Human A3G (hA3G) and A3F (hA3F) proteins inhibit infection by HIV-1 lacking the viral infectivity factor (*vif*) gene^{1,3}. In susceptible cells, hA3G and hA3F are incorporated into *vif*-deficient virions; on infection of target cells, the packaged hA3 proteins deaminate deoxycytidine (C) to deoxyuridine (U) residues in the minus-strand DNA during reverse transcription. By editing C to U, hA3G and hA3F induce G→A

transitions in newly synthesized HIV-1 plus-strand DNA, thus inhibiting viral replication and propagation. Additionally, recent studies indicate that hA3G and hA3F may inhibit viral replication by as yet undefined means that are independent of their deaminase activities⁴.

In contrast with humans, the mouse genome encodes only a single gene for A3. It is not known whether mouse A3 (mA3) protein functions in antiviral immunity to murine retroviruses. Several groups have reported that only hA3G, not mA3, is packaged into murine leukaemia virus (MLV) particles and inhibits their infectivity^{2,5–8}. Moreover, there is evidence that mA3 has a function in restricting the retrotransposition of endogenous mobile genetic elements⁹.

In this context, we examined whether A3 proteins inhibited infection by MMTV, using both *in vitro* and *in vivo* assays. First, we determined whether A3 proteins were packaged into MMTV virions. We transfected an MMTV-producer cell line, CGRES6, with expression vectors encoding V5 epitope-tagged mA3–glutathione S-transferase (GST) or hA3G–GST fusion proteins; as a control, these cells were also transfected with a GST expression vector. Purified virions produced by these cells were analysed by SDS–polyacrylamide-gel electrophoresis (PAGE) and western blotting. Both mA3–GST and hA3G–GST chimaeras (Fig. 1a, compare lanes 6 and 7 with lane 8 in the top two panels) were incorporated into viral particles, and their incorporation had no effect on MMTV virion production (Fig. 1a, compare lanes 6 and 7 with lanes 8 and 9 in the bottom panel). This association of A3

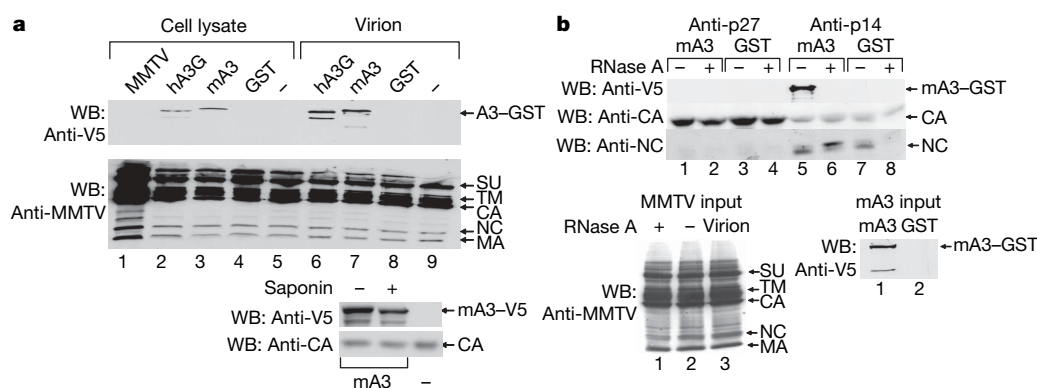


Figure 1 | mA3 and hA3G are incorporated into MMTV virions and bind NC. **a**, Cell lysates and virions purified from hA3G–GST-transfected, mA3–GST-transfected and GST-transfected CGRES6 virus producer cells analysed by western blotting (WB) with anti-V5 and anti-MMTV antibodies. Lane 1, purified MMTV virus; minus signs, extracts (lane 5) and virions (lane 9) from untransfected cells. Smaller fragments represent degradation products. Bottom: an aliquot of mA3-containing virions treated with 0.5% saponin before centrifugation over sucrose cushions. SU, MMTV surface protein

gp52; TM, MMTV transmembrane protein gp36; MA, MMTV matrix protein p10. **b**, Pull-down assays with proteins from disrupted virions immobilized on Protein G–agarose by anti-CA (anti-p27) or anti-NC (anti-p14) monospecific antisera and lysates prepared from 293T cells expressing mA3–GST (mA3) or GST. The anti-NC antiserum has some anti-CA reactivity. Input virion proteins were analysed by western blotting with polyclonal anti-MMTV antisera; input lysates from transfected 293T cells were analysed with anti-V5 antibodies.

¹Department of Microbiology and Abramson Family Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6142, USA. ²Department of Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, SI-1000 Ljubljana, Slovenia. ³Rosalind Russell Medical Research Center, Departments of Medicine, Microbiology and Immunology, University of California at San Francisco, San Francisco, California 94143-0703, USA.

with MMTV virions was not affected by treatment with 0.5% saponin, indicating that this protein was incorporated into viral particles (Fig. 1a).

Retroviral nucleocapsid (NC) proteins have been implicated in the incorporation of A3 into viral particles, although whether this is due to direct protein–protein interactions or to binding of both proteins to viral RNA is unclear¹⁰. We therefore wished to determine whether mA3 also interacts with the MMTV NC in the presence of RNA. Disrupted MMTV virions treated with or without RNase A were immunoprecipitated with anti-MMTV NC (p14) or capsid (CA; p27) antibodies bound to Protein G–agarose beads and incubated with lysates from cells expressing GST or mA3–GST. Bound proteins were eluted from the beads and subjected to western blotting with antibodies against V5 epitope-tagged mA3 and MMTV proteins. As shown in Fig. 1b, the mA3 protein bound to beads containing the MMTV NC but not those containing CA only (compare lanes 5 and 1) in the presence of RNA (compare lanes 5 and 6). These data demonstrate that, unlike MLV, MMTV virions are able to package mA3.

Because A3 proteins were incorporated into virions, we next determined whether they affected viral infectivity. CGRES6 cells produce virions derived from a molecular clone of MMTV with a *gfp* gene (encoding green fluorescent protein) inserted into its long terminal repeat (LTR) called pGR102ES¹¹. Purified virions from CGRES6 cells (Fig. 1) or 293T cells, which coexpressed pGR102ES and V5 epitope-tagged versions of mA3 and hA3G proteins, were used to infect an MMTV-susceptible cell line, TRH3 (293T cells stably expressing the MMTV entry receptor, the mouse transferrin receptor¹²). At 48 h after the infection, we analysed cells for integrated proviruses by fluorescence-activated cell sorting (FACS; Fig. 2a), fluorescence microscopy (Fig. 2b) and polymerase chain reaction (PCR) of genomic DNA (Fig. 2c). In all three assays, both A3 proteins greatly inhibited MMTV infection.

Next, we determined whether the endogenous mA3 protein affected MMTV infection *in vivo*. MMTV infects lymphoid tissues and mammary, salivary and prostate epithelial cells *in vivo*¹³. If mA3

proteins were able to inhibit MMTV infection *in vivo*, they should be expressed in these target tissues. We used semiquantitative reverse-transcriptase-mediated PCR to analyse mA3 expression in lymphoid and non-lymphoid tissues. Target tissues for MMTV all contained mA3 transcripts, although lymph nodes and thymus had higher levels of expression than spleen and epithelial tissues (mammary, salivary and prostate glands; Supplementary Fig. S1). This finding extends the results of a previous study that found mA3 RNA in spleen and thymus¹⁴ and confirms SymAtlas data (<http://symatlas.gnf.org/SymAtlas/>)¹⁵.

To examine whether mA3 affects MMTV infection *in vivo*, we used mA3^{−/−} mice generated by gene trap technology (<http://baygenomics.ucsf.edu/>). In these mice the gene trap in intron 4 leads to the formation of a fusion protein from the first four mA3 exons (exon 1 to exon 4; 197 residues) and β -galactosidase (Supplementary Fig. S2). The absence of deaminase activity of the amino-terminal mutant mA3(1–197) protein was confirmed by an *in vitro* deaminase assay¹⁶ (Supplementary Fig. S3). Moreover, when the mutant mA3(1–197) protein was packaged into HIV particles, it did not inhibit the infection (data not shown). The XN450 embryonic stem cell line was used to generate homozygous mice lacking the wild-type mA3 RNA, as judged by real-time PCR (Supplementary Fig. S5). Indeed, no mutant mA3(1–197)– β -galactosidase fusion protein could be detected in tissues from these mice (data not shown). Homozygous mA3^{−/−} mice had no obvious phenotype, as reported previously for mA3^{−/−} mice generated by conventional gene-targeting approaches¹⁴.

Although mammary epithelial cells are the ultimate targets for MMTV, lymphocytes have a crucial function in the infection *in vivo*¹⁷. In addition to its *gag*, *pol* and *env* genes, MMTV encodes a superantigen (Sag) in its 3' LTR that is expressed only after proviral integration into the host genome. The Sag is presented by major histocompatibility complex (MHC) class II determinants expressed on the surface of infected antigen-presenting cells to T cells bearing specific V β chains of the T-cell antigen receptor. These Sag-specific T cells proliferate, provide B-cell help and produce cytokines that stimulate and recruit lymphoid cells, resulting in a reservoir of infected lymphocytes that carry the virus to the mammary gland¹⁷.

To determine whether lymphocyte infection was affected by A3, adult mA3^{−/−}, mA3^{+/-} and C57BL/6 control mice (five weeks of age) were inoculated subcutaneously with MMTV(RIII), which infects mice with the H-2b major histocompatibility complex locus¹⁸. At four and six days after inoculation, draining and contralateral non-draining lymph nodes were harvested and assayed for evidence of the infection. Two assays were used. First, we examined the infection-dependent, viral Sag-mediated activation of B220⁺ B and CD4⁺ T cells by staining for the CD69 activation marker, and for increases in the percentage of Sag-cognate V β 2⁺CD4⁺ T cells. Second, we used semiquantitative and real-time PCR approaches to examine levels of proviral DNA integration.

Four days after infection, the draining lymph nodes of mA3^{−/−} mice had twofold higher levels of B-cell and T-cell activation and significantly more V β 2⁺, Sag-specific T cells than their wild-type counterparts (Fig. 3); heterozygous mA3^{+/-} mice showed similar activation to wild-type C57BL/6 mice (Supplementary Fig. S6a). Moreover, no activation of non-cognate V β 10⁺/CD4⁺ T cells was detected (Supplementary Fig. S6b). Similar results were obtained on day 6 (data not shown). In addition, at four days after inoculation, viral DNA was detected in the draining lymph nodes of eight out of ten mA3^{−/−} but no wild-type mice (Fig. 4a). Real-time PCR analyses of DNA isolated from these lymph nodes demonstrated more than tenfold greater levels of integrated proviral DNA in the mA3^{−/−} mice than in the wild-type mice (Fig. 4b).

Significantly, the spread of virus was more rapid and extensive in mA3^{−/−} mice. First, both B and T lymphocytes in the non-draining contralateral lymph nodes of the mA3^{−/−} mice but not the wild-type mice were activated at four days after inoculation (compare the

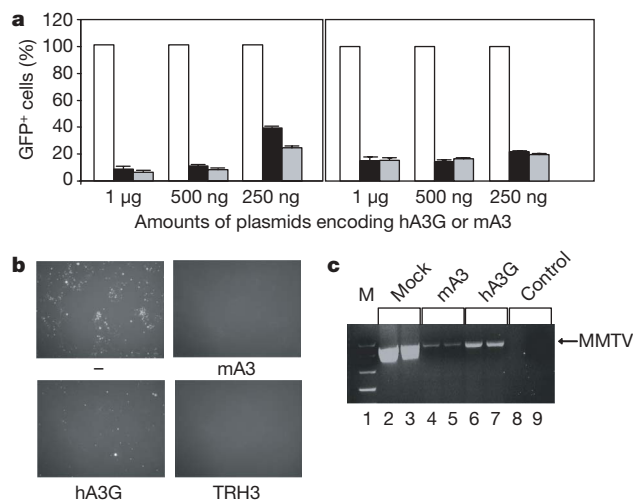


Figure 2 | mA3 and hA3G inhibit MMTV replication *in vitro*. **a**, FACS analyses of TRH3 cells infected with virions containing mouse A3 (grey bars) or human A3G (black bars) prepared from CGRES6 cells (western blots of virus preparations shown in Fig. 1) (left) or with virions prepared from pGR102ES-transfected 293T cells (right). Results are expressed as percentages of GFP-positive cells normalized to virions prepared without A3 proteins (white bars). Error bars indicate s.d. **b**, Fluorescence microscopy of cells infected with virus prepared in CGRES6 cells. Minus sign, virions made in the absence of A3 expression vectors; TRH3, uninfected cells. **c**, Semiquantitative PCR of DNA isolated from infected cells with the use of primers specific to MMTV. Each infection was performed in duplicate. Labels: M, molecular mass markers; mock, virus from mock-transfected cells; control, no DNA.

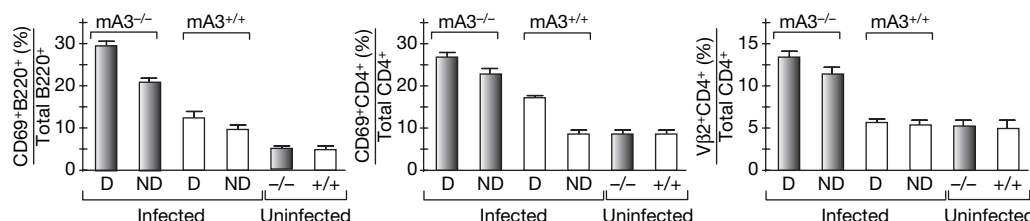


Figure 3 | mA3 inhibits MMTV infection *in vivo*. Four days after inoculation, mice were killed and lymph node cells were harvested, stained with phycoerythrin (PE)-conjugated anti-CD69 and fluorescein isothiocyanate (FITC)-conjugated anti-B220, or with PE-conjugated anti-CD69, allophycocyanin-conjugated anti-CD4 and FITC-conjugated anti-Vβ2 antibodies. Results are presented as percentages calculated from CD69+B220⁺/total B220⁺ B-cell population (left); CD69+CD4⁺/total

CD4⁺ T-cell population (middle) and Vβ2⁺CD4⁺/total CD4⁺ T cells (right). Error bars show s.d.; *n* = 10 mice for each group. Similar results were obtained in two independent experiments. D, draining lymph nodes from inoculated mice; ND, contralateral non-draining lymph nodes from inoculated mice; -/-, uninfected mA3^{-/-} mice; +/+, uninfected wild-type mice.

non-draining lymph nodes from mA3^{-/-} and wild-type mice in Fig. 3). Indeed, B and T cells in the non-draining lymph nodes of mA3^{-/-} mice were more activated than those in the draining lymph nodes of wild-type mice. Moreover, at day 6, levels of viral DNA were fourfold higher in the non-draining lymph nodes and sevenfold higher in the spleens of mA3^{-/-} mice than in those of wild-type mice (Fig. 4b). These higher levels of infection in mA3^{-/-} mice persisted at all sites at 18 days after infection (Fig. 4b). Although at day 18 after inoculation, infection in the draining lymph nodes of mA3^{-/-} mice had increased relative to that at day 4, infection in the periphery (spleen and non-draining lymph nodes) had reached a plateau. In sharp contrast, at the same time, infection in the wild-type mice was still increasing at all sites (Fig. 4b). Heterozygous mA3^{+/-} mice showed infection levels similar to those of wild-type mice at day 18 (data not shown). No infection of the liver, which is not a target tissue, was observed in any of the mice (Fig. 4b). Taken together, these data indicate that mA3 limits MMTV infection *in vivo*.

Last, we determined whether mA3 induces G→A transitions in MMTV proviral DNA. DNA from the spleens of infected wild-type, mA3^{+/-} and mA3^{-/-} mice was amplified by PCR with primers specific to MMTV(RIII). Amplicons were cloned, sequenced and

analysed for mutations. No significant G→A hypermutation was observed (Supplementary Fig. S7). Although the antiretroviral activity of hA3G on Vif-deficient HIV-1 is due in part to the deamination of C to U on the minus strand DNA during reverse transcription², in primary human T cells, which also restrict HIV-1, hA3G was found to cause little or no deamination¹⁹. Indeed, when hA3G with a Q259E mutation that inactivates its deaminase activity was incorporated into HIV or MMTV particles, it retained the antiviral activity in infectivity assays *in vitro* (Supplementary Fig. S4), as reported previously for HIV²⁰. Thus, our data indicate that the mechanism of inhibition by mA3 is not due to its enzymatic activity on the MMTV genome. Alternatively, viruses that were amplified might have survived *in vivo* because they escaped deamination.

Our work describes a significant antiviral function for mA3 in the mouse. Many studies *in vitro* have shown that A3 proteins confer cellular resistance to several exogenous and endogenous viruses. Although previous work with Vif-deficient simian immunodeficiency virus had indirectly implicated APOBEC proteins in attenuating infection in rhesus macaques²¹, our study demonstrates directly that A3 restriction has a function in viral infection *in vivo*. Specifically, MMTV, a milk-borne betaretrovirus that is present in inbred and outbred mice as an exogenous virus, was highly sensitive to the restriction by mA3 *in vitro* and *in vivo*. This finding contrasts with what occurs with other retroviruses such as MLV and MPMV, which prevent the incorporation of A3 proteins of their normal host species^{22,23}.

How, then, do mice become infected with MMTV in the presence of the endogenous mA3 protein? We speculate that mA3 might not totally eliminate MMTV infection but instead partly restricts it. It has been estimated that betaretroviruses entered the mouse genome about 20 Myr ago, just after their speciation^{23,24}. Because MMTV has coexisted with its host almost from the beginning, it must have evolved to avoid the host's antiviral defences. Similarly, these hosts have developed mechanisms for counteracting the pathogenic effects of viral infection. MMTV causes breast cancer in mice with high frequency (more than 90%) but with long latency (more than six months) and therefore does not have a major impact on the viability and reproduction of the species, at least in the laboratory setting¹³. It is possible that this attenuated pathogenesis is due, at least in part, to innate antiviral factors such as mA3. Like HIV-1, the MMTV genome is purine-rich²⁵, indicating that mA3 may have had a role in viral evolution and supporting our hypothesis that the MMTV that persists in mice has become, in an evolutionary sense, less susceptible to inhibition by this protein. Whether the accelerated infection and spread of MMTV in mA3^{-/-} mice will have effects on the ability of the virus to spread and cause breast cancer as well as other pathologies is currently under investigation.

METHODS

Cell lines and plasmid constructions. All cells were cultured in DMEM medium with 10% FCS except for CGRES6-GFP cells¹¹, which were cultured with 10%

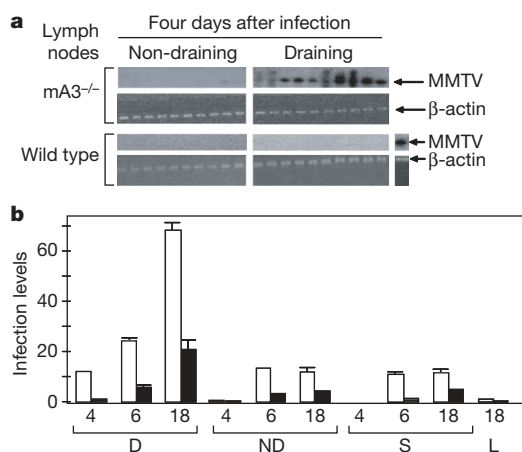


Figure 4 | mA3 inhibits virus infection and spread *in vivo*. **a**, Lymphocyte DNA four days after inoculation was subjected to semiquantitative PCR analysis with MMTV(RIII)-specific primers. After transfer to nitrocellulose, blots were hybridized with an MMTV LTR-specific probe²⁹. In parallel PCR reactions, DNA samples were amplified with mouse β-actin-specific primers. MMTV, DNA isolated from a persistently infected C57BL/6 mouse. **b**, DNAs extracted from draining (D) and non-draining (ND) lymph nodes and spleen (S) at 4, 6 and 18 days, and from liver (L) at 18 days after inoculation, were analysed for MMTV(RIII) sequences by real-time PCR. Open bars, mA3^{-/-}; filled bars, wild type. MMTV signals were normalized to a single-copy mouse gene, *Mttr2*. Error bars show s.d.; *n* = 10 mice (days 4 and 6) and *n* = 5 mice (day 18).

horse serum and antibiotics. The V5 epitope-tagged, GST-tagged mA3 and hA3G plasmid constructions were described previously³.

MMTV packaging and infectivity assays. Transfections were performed with Lipofectamine (Invitrogen). For virions produced in CGRES6–GFP cells, cells were transfected with mA3–GST, hA3G–GST or GST expression vectors. For MMTV virions produced in 293T cells, cells were co-transfected with a GFP-tagged molecular MMTV clone, pGR102ES¹¹, a plasmid containing the rat glucocorticoid receptor and various amounts of mA3–V5, hA3G–V5 or GST expression vectors, as indicated in figure legends. At 24 h after transfection, virus production was induced with 0.5 µM dexamethasone. At 24 h after induction, the virus was purified from cell supernatants by ultracentrifugation on 30% sucrose cushions; in some experiments, viruses were preincubated for 15 min with 0.5% saponin before centrifugation. Resuspended virus pellets were treated with DNase I (Roche). Purified virions were analysed by SDS–PAGE and western blotting and used to infect MMTV-susceptible TRH3 cells. Western blots were probed with rabbit anti-GST (Sigma-Aldrich), anti-V5 (Invitrogen) or anti-MMTV (polyclonal goat anti-MMTV) and appropriate horseradish peroxidase-conjugated secondary antibodies, followed by detection with enhanced chemiluminescence reagents (Amersham Biosciences). Infection assays were performed in six-well plates seeded with TRH3 cells. At 48 h after infection, cells were harvested and assayed for GFP expression by fluorescence microscopy and flow cytometry (FACSCalibur; Becton Dickinson); data were analysed with CellQuest software (Becton Dickinson Immunocytometry Systems). An aliquot of cells was used for DNA isolation (DNAeasy; Qiagen) and subjected to real-time PCR (ABI Prism model 7900HT; Applied Biosystems) or semiquantitative PCR with primers specific for the MMTV LTR²⁶ to detect integrated proviruses. Infectivity assays were performed in triplicate; each assay was performed three times with similar results.

GST pull-down assays. MMTV virions were disrupted with virion lysis buffer (25 mM Tris–HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 0.5% Nonidet P40, 0.5% deoxycholic acid) and incubated for 30 min in the presence or absence of RNase A (50 µg ml^{−1}) at 37 °C. Disrupted virions were immunoprecipitated with goat anti-p27 or goat anti-p14 antisera (National Cancer Institute Biochemical Carcinogenesis Branch Repository) and Protein G (Invitrogen). Lysates prepared from 293T cells transfected with the mA3–GST or GST expression vectors were incubated with the bound virus proteins, eluted with loading buffer and subjected to SDS–PAGE and western blotting.

In vivo MMTV infection. *In vivo* MMTV footprint infectivity assays were performed with virions purified from MMTV(RII) milk, as described previously²⁷. Lymphocytes were isolated from the draining and non-draining lymph nodes and stained with fluorochrome-conjugated monoclonal antibodies against CD4, B220, CD69, Vβ10TCR and Vβ2TCR (BD Biosciences) and subjected to FACS analyses. To examine MMTV infection, DNA isolated from the draining and contralateral non-draining lymph nodes was subjected to PCR with primers specific for the MMTV(RII) LTR; primer sequences are available from the authors on request. After transfer to nitrocellulose, blots were hybridized with an MMTV LTR probe and subjected to autoradiography. Mouse β-actin sequences were amplified from all samples and used as a control for DNA integrity. Levels of integrated MMTV(RII) DNA in infected mouse tissues were also determined by SyBrGreen real-time PCR performed with primers specific for the MMTV(RII) LTR and for a single-copy mouse gene *Mtvr2* (ref. 28). Reactions were performed in triplicate with SyBrGreen 1 master mix and run on an ABI Prism model 7900HT. The reaction conditions were as follows: 10 min at 95 °C, 15 s at 95 °C and 1 min at 60 °C, for 40 cycles. Standard curves were generated for each primer pair by serial dilution of DNA standard templates, and the relative amplification for each sample was quantified from the curve. Data are presented as relative levels of MMTV normalized to the single-copy *Mtvr2* gene.

Received 30 October; accepted 19 December 2006.

Published online 28 January 2007.

1. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
2. Mangeat, B. *et al.* Broad antiretroviral defense by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99–103 (2003).
3. Zheng, Y. H. *et al.* Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *J. Virol.* **78**, 6073–6076 (2004).
4. Holmes, R. K., Koning, F. A., Bishop, K. N. & Malim, M. H. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation: Comparisons with APOBEC3G. *J. Biol. Chem.* advance online publication, doi:10.1074/jbc.M607298200 (22 November 2006)

5. Harris, R. S. *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803–809 (2003).
6. Bishop, K. N. *et al.* Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr. Biol.* **14**, 1392–1396 (2004).
7. Doehle, B. P., Schafer, A., Wiegand, H. L., Bogerd, H. P. & Cullen, B. R. Differential sensitivity of murine leukemia virus to APOBEC3-mediated inhibition is governed by virion exclusion. *J. Virol.* **79**, 8201–8207 (2005).
8. Kobayashi, M. *et al.* APOBEC3G targets specific virus species. *J. Virol.* **78**, 8238–8244 (2004).
9. Esnault, C. *et al.* APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* **433**, 430–433 (2005).
10. Cen, S. *et al.* The interaction between HIV-1 Gag and APOBEC3G. *J. Biol. Chem.* **279**, 33177–33184 (2004).
11. Indik, S., Gunzburg, W. H., Salmons, B. & Rouault, F. Mouse mammary tumor virus infects human cells. *Cancer Res.* **65**, 6651–6659 (2005).
12. Zhang, Y., Rassa, J. C., deObaldia, E. M., Albritton, L. & Ross, S. R. Identification of the mouse mammary tumor virus envelope receptor-binding domain. *J. Virol.* **77**, 10468–10478 (2003).
13. Nandi, S. & McGrath, C. M. Mammary neoplasia in mice. *Adv. Cancer Res.* **17**, 353–414 (1973).
14. Mikl, M. C. *et al.* Mice deficient in APOBEC2 and APOBEC3. *Mol. Cell. Biol.* **25**, 7270–7277 (2005).
15. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101**, 6062–6067 (2004).
16. Svarovskaia, E. S. *et al.* Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. *J. Biol. Chem.* **279**, 35822–35828 (2004).
17. Ross, S. R. Using genetics to probe host–virus interactions: the mouse mammary tumor virus model. *Microbes Infect.* **2**, 1215–1223 (2000).
18. Uz-Zaman, T., Ignatowicz, L. & Sarkar, N. H. Mouse mammary tumor viruses expressed by RII/Sa mice with a high incidence of mammary tumors interact with the Vβ-2- and Vβ-8-specific T cells during viral infection. *Virology* **314**, 294–304 (2003).
19. Chiu, Y. L. *et al.* Cellular APOBEC3G restricts HIV-1 infection in resting CD4⁺ T cells. *Nature* **435**, 108–114 (2005).
20. Newman, E. N. *et al.* Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr. Biol.* **15**, 166–170 (2005).
21. Cullen, B. R. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. *J. Virol.* **80**, 1067–1076 (2006).
22. Doehle, B. P. *et al.* The betaretrovirus Mason–Pfizer Monkey Virus selectively excludes simian APOBEC3G from virion particles. *J. Virol.* **80**, 12102–12108 (2006).
23. Baillie, G. J., van de Lagemaat, L. N., Baust, C. & Mager, D. L. Multiple groups of endogenous betaretroviruses in mice, rats and other mammals. *J. Virol.* **78**, 5784–5798 (2004).
24. Morris, V. L., Medeiros, E., Ringold, G. M., Bishop, J. M. & Varmus, H. E. Comparison of mouse mammary tumor virus-specific DNA in inbred, wild and Asian mice, and in tumors and normal organs from inbred mice. *J. Mol. Biol.* **114**, 73–91 (1977).
25. Berkhout, B., Grigoriev, A., Bakker, M. & Lukashov, V. V. Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. *AIDS Res. Hum. Retroviruses* **18**, 133–141 (2002).
26. Golovkina, T. V., Dudley, J. P., Jaffe, A. & Ross, S. R. Mouse mammary tumor viruses with functional superantigen genes are selected during *in vivo* infection. *Proc. Natl Acad. Sci. USA* **92**, 4828–4832 (1995).
27. Rassa, J. C., Meyers, J. L., Zhang, Y., Kudaravalli, R. & Ross, S. R. Murine retroviruses activate B cells via interaction with Toll-like receptor 4. *Proc. Natl Acad. Sci. USA* **99**, 2281–2286 (2002).
28. Golovkina, T. V. *et al.* A novel membrane protein is a mouse mammary tumor virus receptor. *J. Virol.* **72**, 3066–3071 (1998).
29. Dzuris, J. L., Golovkina, T. V. & Ross, S. R. Both T and B cells shed infectious MMTV. *J. Virol.* **71**, 6044–6048 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The CGRES6 cells and pGR102ES plasmid were a gift from F. Roualt and W. Gunzburg. This study was supported by grants from the National Institutes of Health to S.R.R. and M.B.P. C.O. was supported by a training grant from the National Cancer Institute of the National Institutes of Health and N.L. was funded in part by the Slovenian Research Agency.

Author Contributions C.M.O., N.L., B.M.P. and S.R.R. designed the research. C.M.O. and N.L. performed the research. C.M.O., N.L., B.M.P. and S.R.R. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.R. (rosss@mail.med.upenn.edu).

Foxp3 occupancy and regulation of key target genes during T-cell stimulation

Alexander Marson^{1,2*}, Karsten Kretschmer^{6,7*}, Garrett M. Frampton^{1,2}, Elizabeth S. Jacobsen¹, Julia K. Polansky⁶, Kenzie D. MacIsaac³, Stuart S. Levine¹, Ernest Fraenkel^{4,5}, Harald von Boehmer^{6,7} & Richard A. Young^{1,2}

Foxp3⁺CD4⁺CD25⁺ regulatory T (T_{reg}) cells are essential for the prevention of autoimmunity^{1,2}. T_{reg} cells have an attenuated cytokine response to T-cell receptor stimulation, and can suppress the proliferation and effector function of neighbouring T cells^{3,4}. The forkhead transcription factor Foxp3 (forkhead box P3) is selectively expressed in T_{reg} cells, is required for T_{reg} development and function, and is sufficient to induce a T_{reg} phenotype in conventional CD4⁺CD25[−] T cells^{5–8}. Mutations in *Foxp3* cause severe, multi-organ autoimmunity in both human and mouse^{9–11}. FOXP3 can cooperate in a DNA-binding complex with NFAT (nuclear factor of activated T cells) to regulate the transcription of several known target genes¹². However, the global set of genes regulated directly by Foxp3 is not known and consequently, how this transcription factor controls the gene expression programme for T_{reg} function is not understood. Here we identify Foxp3 target genes and report that many of these are key modulators of T-cell activation and function. Remarkably, the predominant, although not exclusive, effect of Foxp3 occupancy is to suppress the activation of target genes on T-cell stimulation. Foxp3 suppression of its targets appears to be crucial for the normal function of T_{reg} cells, because overactive variants of some target genes are known to be associated with autoimmune disease.

We developed a strategy to identify genes whose promoters are bound by Foxp3 and whose expression is dependent on that transcription factor (Fig. 1). To generate two cell lines that are genetically matched except for Foxp3, we transduced a Foxp3[−]CD4⁺ murine T-cell hybridoma with FLAG-tagged Foxp3. This approach was favoured over comparison of *ex vivo* cells, which are heterogeneous with regard to activation status. The lines provided sufficient numbers of homogeneous cells with appropriate controls to facilitate both location analysis and expression analysis. FACS (fluorescence-activated cell sorting) analysis confirmed that Foxp3 is expressed in the hybridoma at levels comparable to those in *ex vivo* CD4⁺CD25⁺ T_{reg} cells (Supplementary Fig. S1). Previous work has shown that conventional CD4⁺ T cells ectopically expressing Foxp3 do not upregulate interleukin 2 (IL2) secretion following T-cell receptor (TCR) dependent stimulation¹³. To confirm that the Foxp3⁺ hybridomas contain functional Foxp3, we assayed Foxp3[−] and Foxp3⁺ cells for IL2 secretion. Indeed, FACS analysis revealed that IL2 secretion is strongly inhibited in phorbol myristate acetate (PMA)/ionomycin stimulated Foxp3⁺ hybridomas compared to stimulated Foxp3[−] hybridomas (Supplementary Fig. S2).

To identify direct targets of Foxp3, DNA sequences occupied by the transcription factor were identified in a replicate set of experiments using chromatin-immunoprecipitation (ChIP) combined

with DNA microarrays. For this purpose, DNA microarrays were used that contain 60-mer oligonucleotide probes covering the region from −8 kilobases (kb) to +2 kb relative to the transcript start sites for approximately 16,000 annotated mouse genes¹⁴. The sites occupied by Foxp3 were identified as peaks of ChIP-enriched DNA that span closely neighbouring probes (Fig. 2). Foxp3 was found to occupy the promoters of 1,119 genes in PMA/ionomycin stimulated hybridomas (Supplementary Tables S1 and S2). The well-characterized Foxp3 target gene^{12,15}, *Il2*, was among the genes occupied by Foxp3 (Fig. 2a). Most of the promoters occupied by Foxp3 in stimulated T cells were also occupied in unstimulated cells (Supplementary Tables S1 and S2, and Supplementary Fig. S3). However, at some promoters Foxp3 binding increased considerably in cells stimulated with PMA/ionomycin (Fig. 2a and Supplementary Fig. S3). Control immunoprecipitation experiments in Foxp3[−] cells, which produced few positive signals, confirmed the specificity of these results (Supplementary Fig. S4). Our confidence in the binding data was further strengthened by the discovery of a DNA sequence motif, which matches the consensus forkhead motif, at the genomic loci that were bound by Foxp3 (Fig. 2b and Supplementary Table S5). This motif distinguishes Foxp3 bound regions from unbound regions tiled on the promoter arrays with a high level of confidence ($P < 10^{-41}$). Instances of this motif are significantly more likely to be conserved in Foxp3 bound regions than in

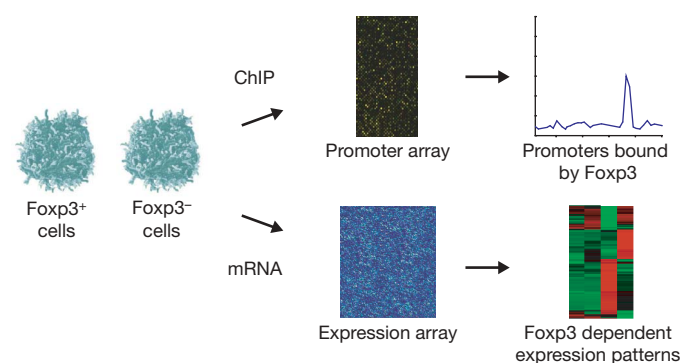


Figure 1 | Strategy to identify direct Foxp3 transcriptional effects. Genetically matched Foxp3⁺ and Foxp3[−] cell populations were generated by transduction of FLAG-tagged Foxp3 into a Foxp3[−] murine T-cell hybridoma. Foxp3 binding sites at promoters across the genome were identified by ChIP experiments with an anti-FLAG antibody. Foxp3 dependent transcriptional regulation was identified by gene expression profiling performed on each of these cell types.

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Department of Biology, ³Department of Electrical Engineering and Computer Science, ⁴Biological Engineering Division, ⁵Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02139, USA. ⁶Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁷Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work

promoter regions that are not bound by Foxp3 ($P < 10^{-23}$), suggesting that these sites serve a functional role (Supplementary Table S5).

To gain insights into the cellular functions that are directly regulated by Foxp3 transcriptional control, we compared the list of genes occupied by Foxp3 in stimulated hybridomas to the biological pathways annotated by the Kyoto Encyclopedia of Genes and Genomes (KEGG)¹⁶. Among these pathways, Foxp3 target genes are most strongly associated with the TCR signalling pathway ($P = 1.4 \times 10^{-5}$) (Fig. 2c). Foxp3 target genes encode proteins that participate at multiple levels of this pathway, including cell surface molecules, signalling components and transcriptional regulators. Many genes with known roles in T cells are missed with existing automated surveys, so we also manually inspected the list of Foxp3 target genes. This

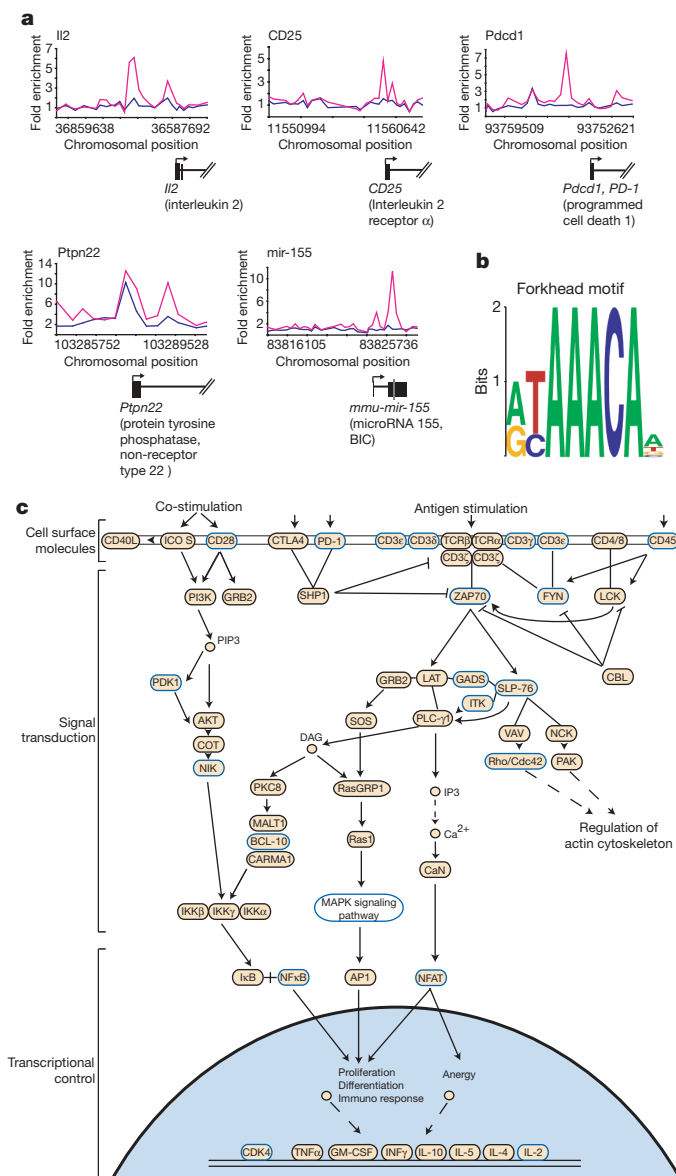


Figure 2 | Direct Foxp3 targets include key modulators of T-cell function.

a, Foxp3 ChIP enrichment ratios (ChIP-enriched versus total genomic DNA) across indicated promoters are shown for stimulated (pink) and unstimulated (blue) cells. Exons (blocks) and introns (lines) of genes and the *mir-155* precursor (grey) are drawn to scale below the plots, with direction of transcription noted by an arrow. **b**, Foxp3 bound genomic regions are enriched for the presence of a forkhead DNA motif, represented here in WebLogo (<http://weblogo.berkeley.edu>). **c**, The KEGG¹⁶ TCR signalling pathway, enriched ($P = 1.4 \times 10^{-5}$) for proteins encoded by direct targets of Foxp3 (blue outline), is displayed.

revealed that many additional Foxp3 targets are probably important for T-cell function, including microRNAs that are differentially expressed between T_{reg} cells and conventional T cells¹⁷ (Fig. 2a and Supplementary Table S3). Surprisingly, *Ctla4* was not among the Foxp3 targets, but analysis by RT-PCR (polymerase chain reaction with reverse transcription) revealed no *Ctla4* expression in Foxp3⁻ and Foxp3⁺ hybridomas (data not shown). However, the Foxp3 targets included genes previously reported to be upregulated in T_{reg} cells, such as *Il2ra* (CD25) (ref. 1), *Tnfrsf18* (GITR) (ref. 18), *Nrp1* (ref. 19) and *Ccr4* (ref. 20), consistent with predictions that these are directly regulated by Foxp3.

Previous reports have shown that only a portion of transcription factor binding events is associated with transcriptional regulation²¹. To identify the set of genes whose expression is dependent on Foxp3, we performed gene expression profiling on Foxp3⁻ and Foxp3⁺ T-cell hybridoma cells before and after PMA/ionomycin stimulation. Comparison of data from unstimulated Foxp3⁻ and Foxp3⁺ cells revealed few differentially expressed genes, suggesting that the transcription factor had little influence on global gene expression in unstimulated hybridomas (Supplementary Table S8). In contrast, PMA/ionomycin-stimulated Foxp3⁻ and Foxp3⁺ cells showed significant differences in expression of almost 1% of mouse genes. Many of these genes were directly occupied by Foxp3, and Foxp3 binding was predominantly, but not exclusively, associated with genes whose expression is downregulated in stimulated Foxp3⁺ hybridomas. Foxp3 occupied the promoters of approximately half of the genes in this cluster (Fig. 3a). This set of downregulated target genes is enriched for genes that are implicated in TCR signalling ($P = 6.1 \times 10^{-3}$). Underscoring the significance of this finding, E2f4, an unrelated control transcription factor that occupies ~800 genes in these hybridomas, was found to occupy only one of these promoters (Supplementary Table S4).

Only a subset of all Foxp3 occupied genes was found to be differentially expressed. One reason for this could be that Foxp3 requires cofactors to modulate transcription. Recently, FOXP3 was shown to cooperate with NFAT in a DNA-binding complex to activate or repress target gene expression¹². Consistent with this report, Foxp3 exerts a more pronounced transcriptional effect in stimulated hybridoma cells than in non-stimulated cells. Importantly, in our experiments, the set of Foxp3 bound genes are enriched ($P < 10^{-19}$) for the presence of an Nfat DNA sequence motif neighbouring the sites of Foxp3 occupancy (Supplementary Table S5).

We next examined whether the genes regulated by Foxp3 in the T-cell hybridomas show similar regulatory behaviour in *ex vivo* T cells. Hybridoma cells were used initially because they afforded cells that differ only by Foxp3 status, they provided a homogenous population of non-stimulated T cells, and we could have confidence in genome-wide location data using FLAG-tagged Foxp3. Nonetheless, microarray expression profiling was performed on *ex vivo* T cells from different mice that express a transgenic TCR alone or together with the TCR agonist ligand²², from which relatively pure populations of naïve CD4⁺CD25⁻ T helper and Foxp3⁺CD4⁺CD25⁺ T_{reg} cells, respectively, could be isolated. This analysis revealed that many, although not all, of the regulated targets that were identified in the hybridomas show consistent expression patterns in stimulated *ex vivo* cells (Fig. 3b and Supplementary Fig. S5). Some differences in gene expression, especially genes that are activated by Foxp3 only in T_{reg} cells, may be due to the fact that *ex vivo* T_{reg} cells are generated through antigenic stimulation^{23,24} and hence could contain transcriptional cofactors that differ from those in hybridoma cells.

Our findings were further validated for a panel of nine Foxp3 targets. Site-specific primers were used to confirm the binding of Foxp3 to the promoters of these genes (Fig. 3c) and quantitative RT-PCR was used to assay messenger RNA levels in non-stimulated and stimulated hybridomas and *ex vivo* cells, in the presence and absence of 2 μ M cyclosporin A (Fig. 3d). These experiments confirmed the direct effects of Foxp3 at targets that were identified in

the genome scale experiments. Notably, in these experiments (Fig. 3d) all genes that were activated following stimulation in Foxp3^- cells and repressed in Foxp3^+ cells were activated in a calcineurin dependent manner, consistent with the notion that Nfat is involved in their activation. Finally, just as Foxp3 regulates the protein levels of secreted IL2 (Supplementary Fig. S2), cell surface

staining and FACS analysis show that Foxp3 regulates the level of Ly6a protein, demonstrating that Foxp3 transcriptional regulation of its targets modulates protein levels (Supplementary Fig. S6).

Taken together, the results from *ex vivo* T cells and the hybridoma system identify a core set of Foxp3 regulatory targets (Fig. 4a), most of which showed suppressed activation in stimulated Foxp3^+ cells. A

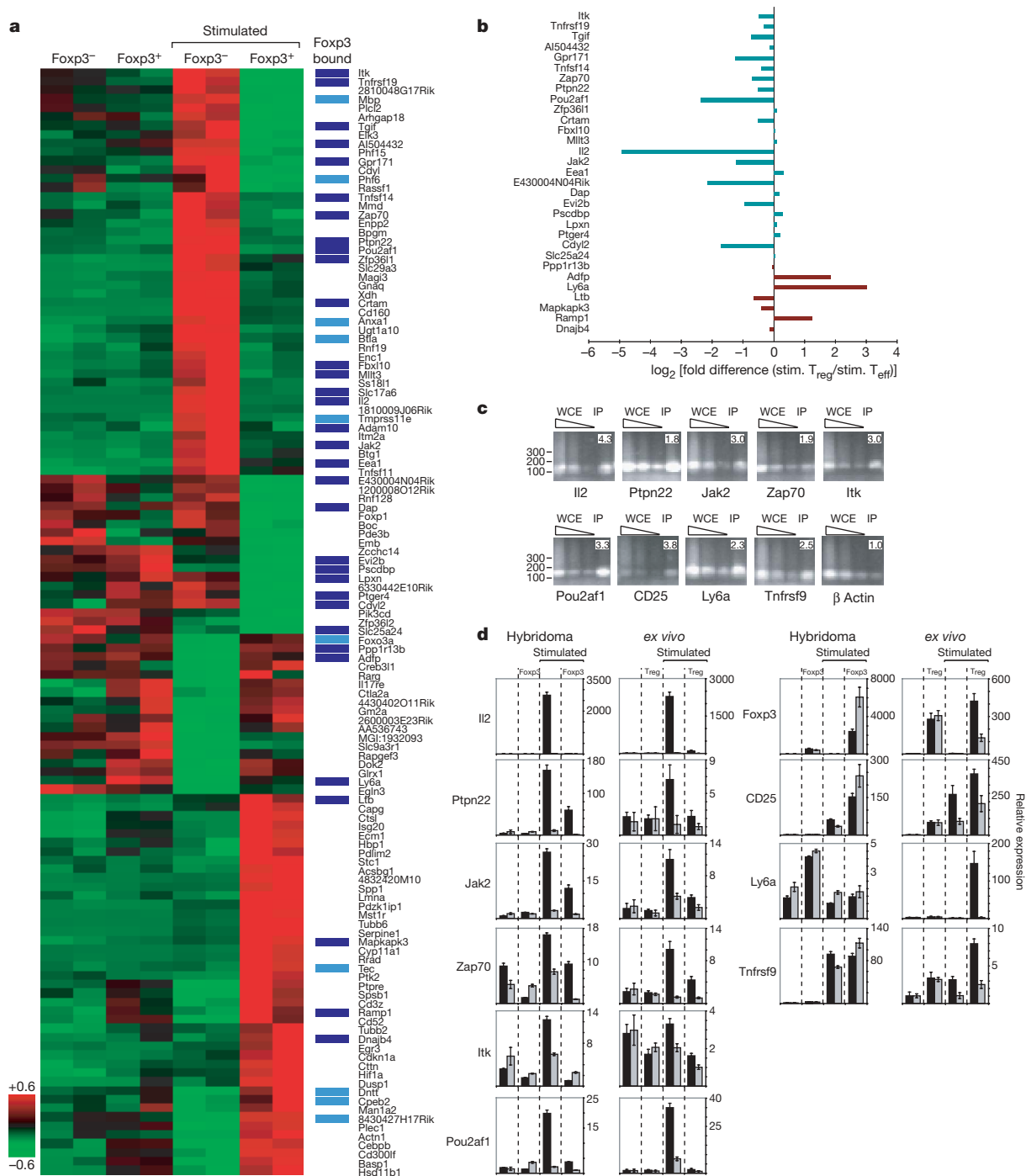


Figure 3 | Foxp3 directly suppresses the activation of target genes. **a**, Replicate expression data for the 125 genes with Foxp3 dependent differential expression in stimulated hybridomas (false discovery rate, FDR < 0.05) were hierarchically clustered and displayed. The Z-score normalized induction (red) or repression (green) is shown for each gene. Direct targets of Foxp3 in stimulated hybridomas are indicated (dark blue for FDR < 0.05, light blue for FDR < 0.10). **b**, For repressed (green) and induced (red) Foxp3-bound targets in **a**, log₂(fold difference) in expression between stimulated *ex vivo* effector (T_{eff}) T cells and T_{reg} cells is displayed. *Slc17a6* and *Adam10* are not expressed in the *ex vivo* samples. **c**, Site-specific

PCR on 10 ng of ChIP DNA confirmed selected targets. Immunoprecipitated (IP) DNA was compared to serial dilutions (90, 30 and 10 ng of DNA) of unenriched whole cell extract (WCE) DNA. Enrichment ratios, shown at top right of each sub-panel, are normalized relative to the unenriched β actin control. DNA fragment size (bp) is indicated on the left of each row. **d**, The transcript levels of the panel of Foxp3 targets presented in **c** and of *Foxp3* were analysed by real-time RT-PCR in stimulated and unstimulated cells, with (grey) and without (black) cyclosporin A. Mean values \pm s.d. of relative expression, determined in triplicate, are shown for indicated genes.

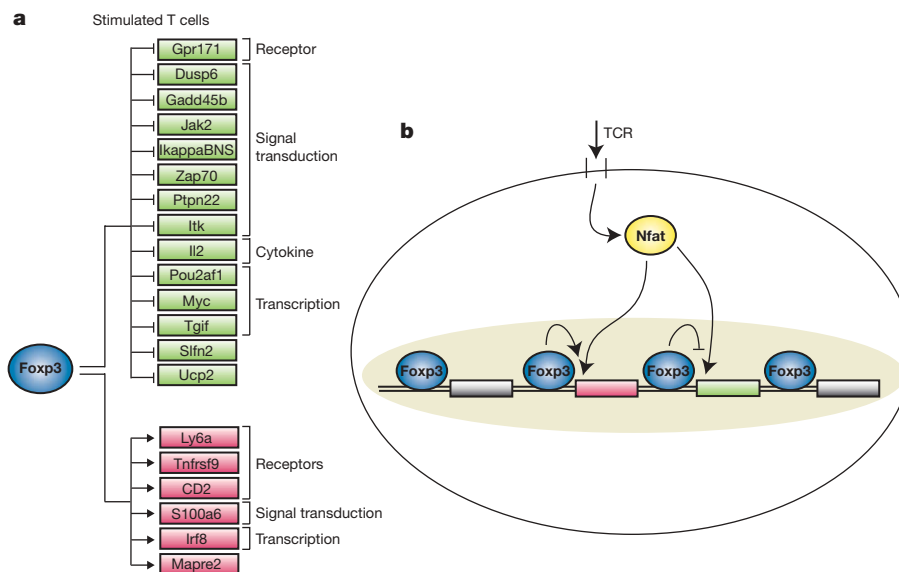


Figure 4 | Core direct regulatory effects of Foxp3. **a**, Shown here are a subset of direct Foxp3 targets that exhibit consistent transcriptional behaviour in hybridomas and in *ex vivo* T cells (Supplementary Fig. S5). **b**, Foxp3 binds to a large set of promoters both in unstimulated and stimulated T cells, but Foxp3 transcriptional regulation is more

extensive in stimulated T cells. The genomic regions where Foxp3 binds are enriched for an Nfat binding site DNA motif. In the hybridomas, Foxp3 predominantly acts to directly suppress the activation of its target genes.

smaller number of Foxp3 target genes was upregulated in stimulated Foxp3⁺ cells, including some encoding cell surface molecules with known roles in immunoregulation, such as Ly6a (ref. 25) and Tnfrsf9 (4-1BB) (ref. 26). The results from the hybridoma system indicate that Foxp3 occupies regions of most of its target promoters in both unstimulated and stimulated conditions, but increased binding at some promoters and regulation of most targets is observed after stimulation. Furthermore, in hybridomas the major function of Foxp3 at these genes is to suppress the level of gene activation that would occur if this transcription factor were not expressed (Fig. 4b). Conceivably, the Foxp3 dependent downregulation of T-cell activation and cytokine genes, and upregulation of immunosuppressive cell surface molecules, contribute to both the hyporesponsive and suppressive T_{reg} phenotype.

Mutations in some Foxp3 target genes are already known to be associated with autoimmune disease. The protein tyrosine phosphatase Ptpn22 is a notable example. In our experiments, *Ptpn22* is one of the highest confidence direct targets of Foxp3, it is upregulated on stimulation in Foxp3⁺ cells, and this upregulation is inhibited in Foxp3⁺ hybridomas (Fig. 3a) and *ex vivo* T_{reg} cells (Fig. 3b, d). *Ptpn22* modulates the signal cascade downstream of the TCR, and mutations in the human *PTPN22* have been associated with type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus and Graves' disease, as well as other autoimmune diseases^{27–29}. A recent report suggests that one *PTPN22* single-nucleotide polymorphism associated with autoimmunity is a gain-of-function mutation³⁰. Our findings are compatible with the hypothesis that the gain-of-function mutation might be pathogenic if mutant *PTPN22* is overactive in T_{reg} cells³⁰.

In summary, our data indicate that Foxp3 binds to the promoters of well-characterized regulators of T-cell activation and function. In the T-cell hybridomas studied here, the major role of this transcription factor is to dampen the induction of key genes when T_{reg} cells are stimulated. In *ex vivo* T_{reg} cells, Foxp3 could also activate the expression of a greater number of genes, perhaps owing to the greater abundance of certain transcriptional cofactors in these cells. Some of the identified Foxp3 target genes have been previously implicated in autoimmune diseases, implying that a therapeutic strategy to recapitulate the function of this transcription factor may have clinical utility for these diseases.

METHODS

A detailed description of all materials and methods used can be found in Supplementary Information.

Growth of murine CD4⁺ T-cell hybridomas and *ex vivo* T cells. CD4⁺ 5B6-2 hybridoma cells expressing a PLP_{139–151}-specific TCR, which was kindly provided by V. Kuchroo, were cultured in Dulbecco's modified Eagle medium (Invitrogen). Primary murine CD4⁺ T cells were cultured in RPMI-1640 medium (Invitrogen). For gene expression profiling, real-time RT-PCR, and location analysis, cells were cultured in the absence or presence of 50 ng ml⁻¹ phorbol 12-myristate 13-acetate (PMA) and 200 ng ml⁻¹ ionomycin at 37 °C and harvested after 6 h. Where indicated, cells were preincubated for 1 h with 2 μM cyclosporin A. Details of cell generation and isolation are provided in Supplementary Information.

Antibodies and ChIP assays. Detailed descriptions of antibodies, antibody specificity and ChIP methods used in this study are provided in Supplementary Information.

Promoter array design and data extraction. The design of the oligonucleotide-based promoter array set and data extraction methods are described in Supplementary Information. The microarrays used for location analysis in this study were manufactured by Agilent Technologies (<http://www.agilent.com>).

Motif analysis. Discovery of the Foxp3 sequence motif from the ChIP-chip binding data was performed using the THEME algorithm. The Foxp3 motif learned by THEME and the Nfat motif from the TRANSFAC database (version 8.3) were used to scan all arrayed sequences to identify matches to the motifs.

Functional classification of bound genes. Comparison of Foxp3 target genes to annotated KEGG biological pathways was performed using the online DAVID tool (<http://niaid.abcc.ncifcrf.gov/>).

Gene expression profiling. For each hybridoma culture condition, total RNA was prepared from 1 × 10⁷ cells using Trizol (Gibco) followed by additional purification using the RNeasy Mini Kit (Qiagen). Biotinylated antisense cRNA was then prepared according to the Affymetrix standard labelling protocol (one amplification round). For each primary T-cell culture condition, total RNA was isolated from 5 × 10⁵ cells with RNeasy. Biotinylated antisense cRNA was prepared by two rounds of *in vitro* amplification using the BioArray RNA Amplification and Labelling System (Enzo Life Sciences) according to the protocol for 10–1,000 ng of input RNA provided by the manufacturer. Biotinylated cRNAs of hybridomas and primary T cells were fragmented and hybridized to Affymetrix GeneChip Mouse Expression Set 430 2.0 arrays at the Microarray Core Facility (Dana-Farber Cancer Institute).

Quantitative RT-PCR. To determine transcript levels in T-cell hybridomas and *ex vivo* T cells, RNA was isolated, reverse-transcribed and subjected to real-time PCR performed on an ABI PRISM thermal cycler using SYBR Green PCR core reagents (Applied Biosystems). Detailed information is provided in Supplementary Information.

Received 13 October; accepted 27 November 2006.

Published online 21 January 2007.

1. Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor α -chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.* **155**, 1151–1164 (1995).
2. Baecher-Allan, C. & Hafler, D. A. Human regulatory T cells and their role in autoimmune disease. *Immunol. Rev.* **212**, 203–216 (2006).
3. Shevach, E. M. CD4⁺ CD25⁺ suppressor T cells: more questions than answers. *Nature Rev. Immunol.* **2**, 389–400 (2002).
4. von Boehmer, H. Mechanisms of suppression by suppressor T cells. *Nature Immunol.* **6**, 338–344 (2005).
5. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061 (2003).
6. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4⁺ CD25⁺ regulatory T cells. *Nature Immunol.* **4**, 330–336 (2003).
7. Khattri, R. *et al.* An essential role for Scurfin in CD4⁺ CD25⁺ T regulatory cells. *Nature Immunol.* **4**, 337–342 (2003).
8. Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341 (2005).
9. Brunkow, M. E. *et al.* Disruption of a new forkhead/winged-helix protein, scurf, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nature Genet.* **27**, 68–73 (2001).
10. Bennett, C. L. *et al.* The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nature Genet.* **27**, 20–21 (2001).
11. Wildin, R. S. *et al.* X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. *Nature Genet.* **27**, 18–20 (2001).
12. Wu, Y. *et al.* FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell* **126**, 375–387 (2006).
13. Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F. & Ziegler, S. F. Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. *J. Biol. Chem.* **276**, 37672–37679 (2001).
14. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
15. Chen, C., Rowell, E. A., Thomas, R. M., Hancock, W. W. & Wells, A. D. Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. *J. Biol. Chem.* **281**, 36828–36834 (2006).
16. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
17. Cobb, B. S. *et al.* A role for Dicer in immune regulation. *J. Exp. Med.* **203**, 2519–2527 (2006).
18. McHugh, R. S. *et al.* CD4⁺ CD25⁺ immunoregulatory T cells: gene expression analysis reveals a functional role for the glucocorticoid-induced TNF receptor. *Immunity* **16**, 311–323 (2002).
19. Bruder, D. *et al.* Neuropilin-1: a surface marker of regulatory T cells. *Eur. J. Immunol.* **34**, 623–630 (2004).
20. Iellem, A. *et al.* Unique chemotactic response profile and specific expression of chemokine receptors CCR4 and CCR8 by CD4⁺ CD25⁺ regulatory T cells. *J. Exp. Med.* **194**, 847–853 (2001).
21. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
22. Klein, L., Khazaie, K. & von Boehmer, H. *In vivo* dynamics of antigen-specific regulatory T cells not predicted from behavior *in vitro*. *Proc. Natl Acad. Sci. USA* **100**, 8886–8892 (2003).
23. Jordan, M. S. *et al.* Thymic selection of CD4⁺ CD25⁺ regulatory T cells induced by an agonist self-peptide. *Nature Immunol.* **2**, 301–306 (2001).
24. Kretschmer, K. *et al.* Inducing and expanding regulatory T cell populations by foreign antigen. *Nature Immunol.* **6**, 1219–1227 (2005).
25. Stanford, W. L. *et al.* Altered proliferative response by T lymphocytes of Ly-6A (Sca-1) null mice. *J. Exp. Med.* **186**, 705–717 (1997).
26. Myers, L. M. & Vella, A. T. Interfacing T-cell effector and regulatory function through CD137 (4-1BB) co-stimulation. *Trends Immunol.* **26**, 440–446 (2005).
27. Bottini, N. *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes. *Nature Genet.* **36**, 337–338 (2004).
28. Wu, J. *et al.* Identification of substrates of human protein-tyrosine phosphatase PTPN22. *J. Biol. Chem.* **281**, 11002–11010 (2006).
29. Bottini, N., Vang, T., Cucca, F. & Mustelin, T. Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin. Immunol.* **18**, 207–213 (2006).
30. Vang, T. *et al.* Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nature Genet.* **37**, 1317–1319 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Young, von Boehmer and Fraenkel laboratories, as well as R. Jaenisch and D. K. Gifford, for discussions and critical review of the manuscript, especially T. I. Lee, J. Zeitlinger and D. T. Odom. We also thank Biology and Research Computing (BaRC), especially T. Dicesare for graphic assistance, as well as E. Herbolzheimer for computational and technical support. K.K. was supported in part by a fellowship grant from the German Research Foundation. This work was supported in part by a donation from E. Radutzky, and by the Whitaker Foundation (E.F.) and the NIH (H.v.B. and R.A.Y.).

Author Contributions R.A.Y. and H.v.B. contributed both as senior and corresponding authors.

Author Information All microarray data from this study are available from ArrayExpress at the EBI (<http://www.ebi.ac.uk/arrayexpress>) under accession code E-TABM-154. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.A.Y. (young@wi.mit.edu) or H.v.B. (Harald_von_Boehmer@dfci.harvard.edu).

LETTERS

Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells

Ye Zheng², Steven Z. Josefowicz², Arnold Kas², Tin-Tin Chu², Marc A. Gavin^{2†} & Alexander Y. Rudensky^{1,2}

Transcription factor Foxp3 (forkhead box P3), restricted in its expression to a specialized regulatory CD4⁺ T-cell subset (T_R) with a dedicated suppressor function, controls T_R lineage development. In humans and mice, Foxp3 deficiency results in a paucity of T_R cells and a fatal breach in immunological tolerance, causing highly aggressive multi-organ autoimmune pathology^{1–3}. Here, through genome-wide analysis combining chromatin immunoprecipitation with mouse genome tiling array profiling, we identify Foxp3 binding regions for ~700 genes and for an intergenically encoded microRNA. We find that a large number of Foxp3-bound genes are up- or downregulated in Foxp3⁺ T cells, suggesting that Foxp3 acts as both a transcriptional activator and repressor. Foxp3-mediated regulation unique to the thymus affects, among others, genes encoding nuclear factors that control gene expression and chromatin remodelling. In contrast, Foxp3 target genes shared by the thymic and peripheral T_R cells encode primarily plasma membrane proteins, as well as cell signalling proteins. Together, our studies suggest that distinct transcriptional sub-programmes implemented by Foxp3 establish T_R lineage during differentiation and its proliferative and functional competence in the periphery.

For the analysis of direct targets of Foxp3, we established a highly specific Foxp3 ChIP (chromatin immunoprecipitation) protocol using the *Pde3b* gene encoding phosphodiesterase 3b as a candidate Foxp3 target repressed by Foxp3 in T_R cells (ref. 4). Foxp3 ChIP was performed using chromatin isolated from a purified CD25⁺CD4⁺ T_R population containing ~90% Foxp3⁺ T cells (see Methods). Foxp3-bound chromatin DNA was probed using primer sets covering 23 conserved non-coding *Pde3b* regions⁵ (Supplementary Fig. 1a). We found that Foxp3 occupies a conserved region within the first intron of *Pde3b* (Supplementary Fig. 1; data not shown). These results highlight the specificity of Foxp3 ChIP and show that *Pde3b* is a direct target of Foxp3-mediated repression, in agreement with the low level of gene and protein expression unique to T_R cells⁴.

For global identification of Foxp3-bound genes, we used the Affymetrix 2.0R mouse genome tiling array. Foxp3 ChIP was performed using the nuclear fraction from freshly isolated CD4⁺CD25⁺ T_R cells from C57BL/6 mice and CD4⁺ T cells from *Foxp3*[−] mice (negative control). To validate array results, quantitative polymerase chain reaction (qPCR) analysis of randomly selected Foxp3 binding peak regions was performed using Foxp3 ChIP DNA. As shown in Fig. 1, binding of Foxp3 to the corresponding regions was independently

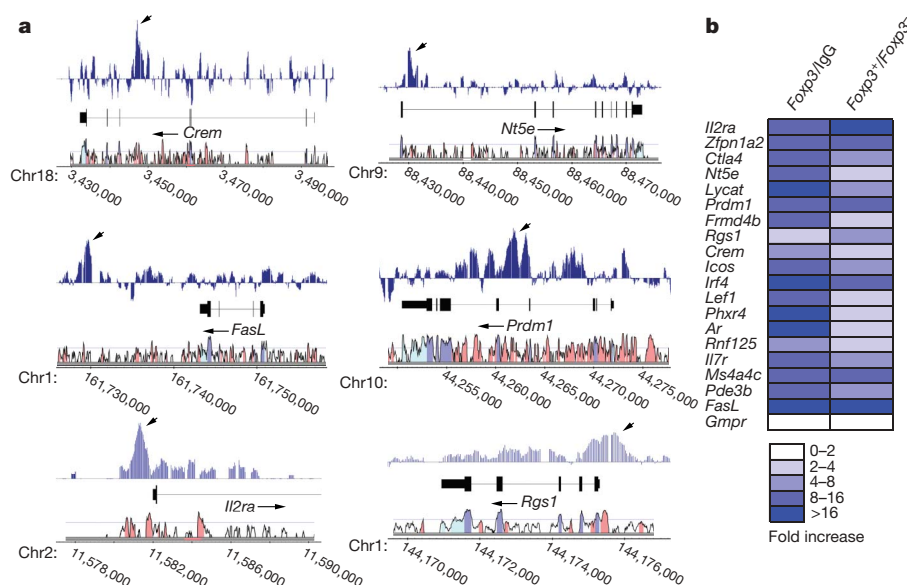


Figure 1 | Analysis of Foxp3-bound genes using a mouse genome tiling array. **a**, Foxp3-binding regions for several genes visualized using the Affymetrix Integrated Genome Browser. Each blue bar represents the signal intensity of an individual oligonucleotide probe. Structure, chromosomal location, and mouse-human homology VISTA plot of the given gene are shown below each peak. **b**, Confirmation of Foxp3 binding to corresponding

peak locations by ChIP-qPCR. The first column shows the comparison between Foxp3 antibody and IgG control precipitated chromatin-bound DNA from B6 T_R cells. The second column shows the comparison between Foxp3 antibody precipitated chromatin-bound DNA isolated from B6 T_R cells and *Foxp3*[−]CD4⁺ T cells. *Gmpr* served as a negative control as this gene is not differentially expressed in all CD4 T-cell subsets.

¹Howard Hughes Medical Institute, ²Department of Immunology, University of Washington, Seattle, Washington 98195, USA. [†]Present address: Amgen Corporation, Seattle, Washington 98101, USA.

confirmed using Foxp3 ChIP-qPCR for 90% of the tested peaks with P values $<10^{-8}$. For Foxp3 binding peaks with P values between 10^{-6} and 10^{-8} , ChIP-qPCR confirmed Foxp3 binding to $\sim 70\%$ of the target regions (data not shown). On the basis of this analysis, we chose a P value of 10^{-6} as the cut-off. Using this threshold, Foxp3 binds to 1,276 regions genome-wide. Computer-assisted mapping of the Foxp3 regions to the locations of known and predicted genes in the mouse genome revealed 702 genes located within 50 kilobases (kb) upstream or 50 kb downstream of Foxp3 binding regions (Supplementary Table 1).

To assess the distribution of Foxp3 binding sites within the genic regions, we compared the positions of Foxp3 binding peaks relative to locations of known and predicted genes. Foxp3 binding sites were substantially enriched within the 0–10 kb upstream regions, the 5' untranslated region (UTR), and the first introns (Fig. 2a, b). However, only a small fraction of Foxp3 binding peaks mapped to the 3'-UTR and coding exons. Further analysis of all the peaks located within 50 kb upstream of a gene showed that the frequency of peaks progressively decreased with increasing distance from the transcription start

site (Fig. 2c). A closer examination of regions 0–3 kb upstream of transcription start sites revealed the highest density of Foxp3 binding sequences within 0–1.2 kb upstream of transcription start sites, corresponding to the promoter regions. Similar analysis of peaks located 0–50 kb downstream of transcription end sites showed that Foxp3 binding sequences occur more frequently within the 0–10 kb downstream of a gene, although the difference is less pronounced compared to the peaks upstream of transcription start sites (Fig. 2c). Preferential positioning of Foxp3 binding sites in the proximity of promoters or within the first introns is evidence for the role of Foxp3 in classical transcriptional regulation of gene expression.

In order to reveal general functional features of the molecular programme implemented by Foxp3, we conducted gene ontology analysis by comparing 702 Foxp3-bound genes with three independently generated lists each containing a similar number of randomly selected genes (Supplementary Table 2). We found Foxp3 binding genes enriched for genes involved in the TCR signalling pathway, cell communication, and transcriptional regulation. These results suggested a role for Foxp3 in setting up transcription factor and receptor

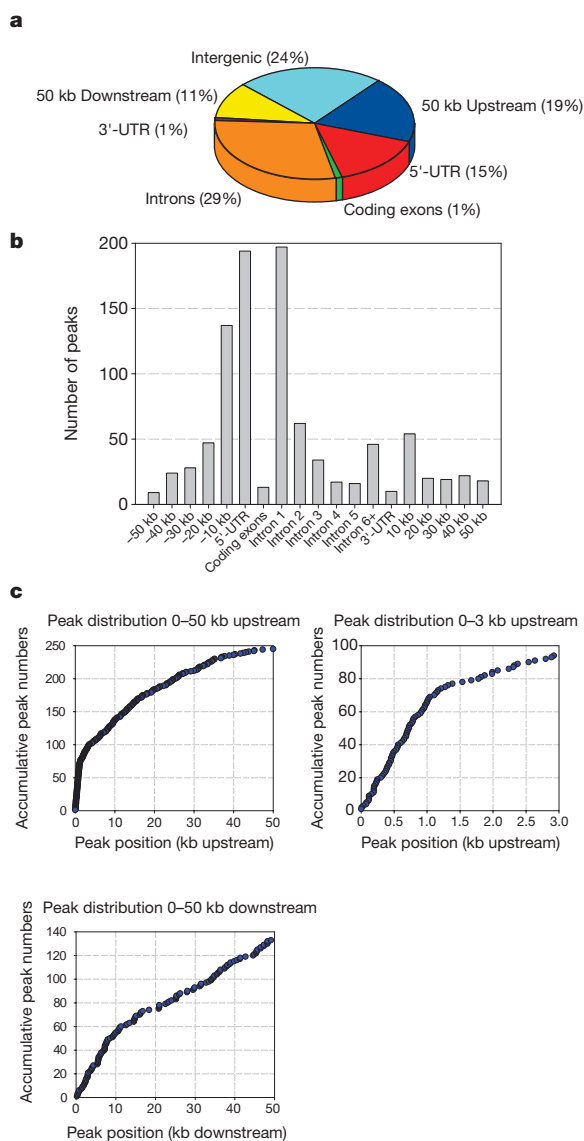


Figure 2 | Distribution of Foxp3-binding regions within the mouse genome. **a**, **b**, Foxp3-binding sequence locations relative to known and predicted genes in the mouse genome. **c**, Foxp3 binding peak distribution within the 50-kb and 3-kb regions upstream of a target gene, or within the 50-kb region downstream of a target gene.

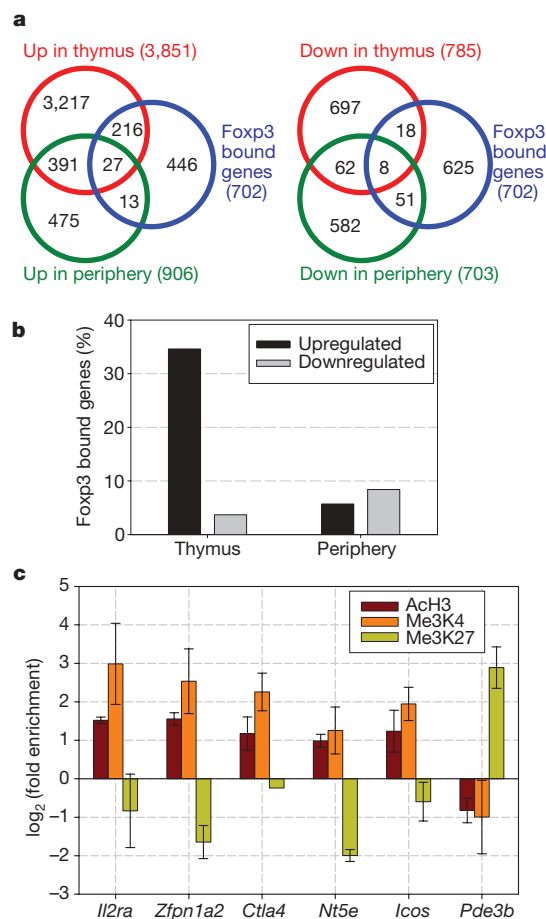


Figure 3 | Foxp3 binding results in both activation and repression of its target genes. A distinct pattern of up- and downregulation is found in the thymus and in the periphery. **a**, Relationship between Foxp3-bound genes and Foxp3-dependent genes revealed by expression array analysis. Genes with expression increased or decreased by twofold or more in T_H versus T_{FN} were considered upregulated or downregulated by Foxp3. **b**, Percentages of Foxp3-bound genes differentially expressed in either thymus or periphery. **c**, Histone modifications of Foxp3-bound regions of Foxp3 target genes. ChIP was performed using *ex vivo* isolated T_H cells and antibodies against acetylated H3K9/14 (AcH3), tri-methyl H3K4 (Me3K4), and tri-methyl H3K27 (Me3K27). For qPCR, primers corresponding to the Foxp3-bound regions of the target genes were used. Identical amounts of precipitated or input DNA were used for each PCR reaction. Results are representative of two experiments; error bars, ± 1 s.d.

signalling networks in T_R cells involved in their development and maintenance.

To explore this possibility, we next identified subsets of Foxp3-bound genes whose expression is altered during T_R cell development in the thymus or in mature T_R cells in the periphery. Recently, we characterized Foxp3-dependent gene expression programme through transcriptional profiling of T_R cells from previously described Foxp3^{3^{sp}} mice and GFP-marked T cells transcribing the Foxp3 gene yet lacking Foxp3 protein expression (T 'Foxp3 null' or T_{FN})^{4,6}. The latter cells were isolated from heterozygote female Foxp3^{3^{sp}/wt} mice, which remained healthy owing to random X-chromosome inactivation⁴. Through comparison of T_R with T_{FN} and 'naive' Foxp3⁻CD25⁻ (T_N) transcriptomes, we generated a list of genes whose expression changed by more than twofold in a Foxp3-dependent manner. By cross-referencing this data set with the set of Foxp3-bound genes, we were able to identify putative Foxp3 direct target genes (Fig. 3a; Supplementary Fig. 3; Supplementary Table 3) and to show that a large number of Foxp3-bound genes are differentially expressed in T_R cells relative to T_{FN} and T_N cells. In contrast, for transcription factors like CREB or T-bet, occupancy of the corresponding binding sites in transformed and primary cells correlated with the changes in expression of very few of these genes^{7,8}.

Upon closer examination, ~35% of the Foxp3-bound genes were upregulated in T_R cells in the thymus and ~6% in the periphery, while a smaller proportion of Foxp3-bound genes were

downregulated either in the thymus or periphery (Fig. 3b). One substantial caveat in this analysis is that limiting numbers of thymic Foxp3⁺ cells prevent direct identification of Foxp3-bound genes in these cells. Therefore, we cannot exclude the existence of additional Foxp3-bound genes unique to the thymus. Nevertheless, our results suggest that Foxp3 mediates distinct transcriptional programmes in the thymus and in the periphery, and can act as both a transcriptional activator and repressor.

This result challenges the prevalent view of Foxp3 as a transcriptional repressor³. A recent study suggested that both repression of *Il2* and activation of *Il2ra* and *Ctla4* transcription occur upon cooperative NFAT and Foxp3 binding to corresponding promoters⁹. Notably, in freshly isolated or cultured T_R cells, we do find Foxp3 bound to the *Il2ra* promoter region but not to the *Il2* gene (Supplementary Fig. 2). Upon stimulation with phorbol myristate acetate (PMA) and ionomycin, we observed an increase in the Foxp3 occupancy of the *Il2* promoter. These results are in full agreement with *Il2* repression that we observed in both T_R and T_{FN} cells, which was, however, more readily reversible in the latter cells upon strong activation⁴. These results are also consistent with the Foxp3 binding to the *Il2* promoter observed in Foxp3-transduced cells only upon provision of a strong stimulus^{3,10}.

To investigate if histone modification is part of the mechanism of Foxp3-mediated gene regulation, we examined permissive and inhibitory histone H3 modifications at Foxp3 binding regions for several

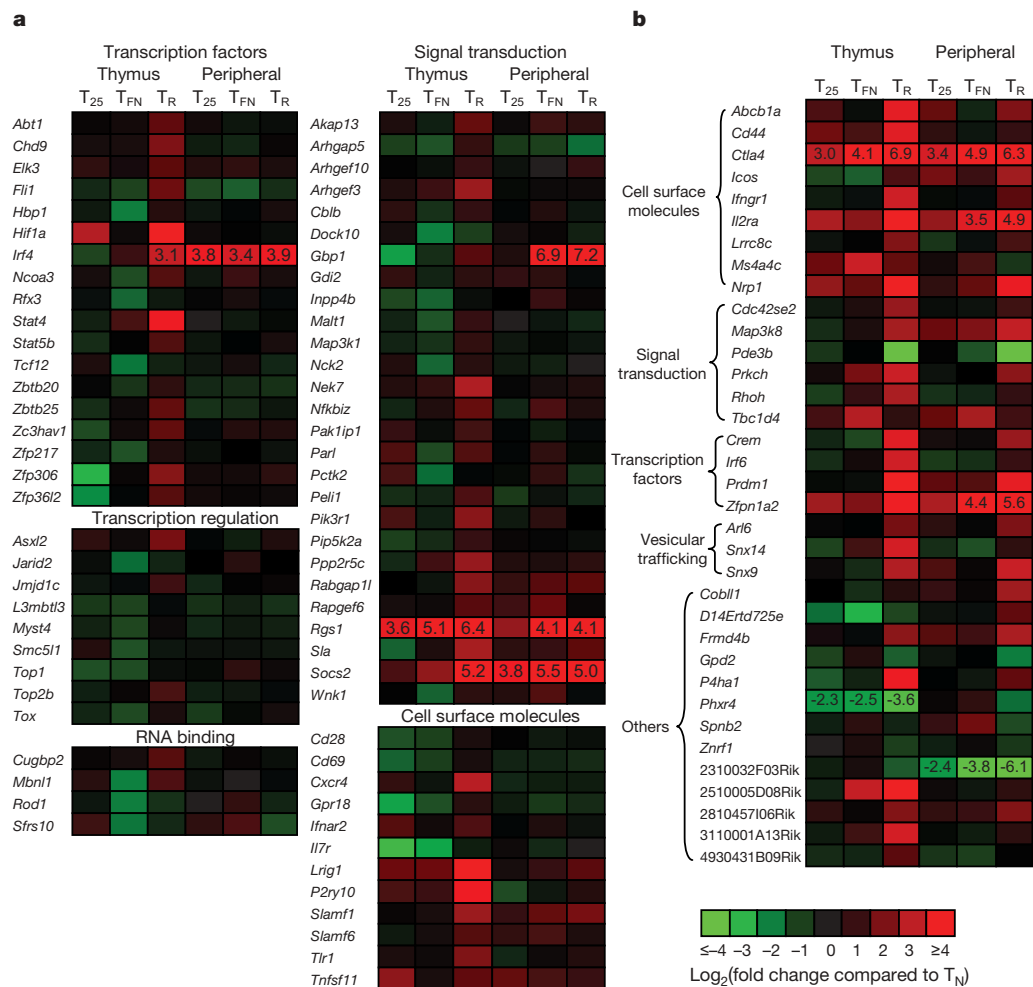


Figure 4 | Differential expression of prevalent functional groups of Foxp3-bound target genes in thymic and peripheral T_R cells. **a**, Foxp3-bound genes differentially expressed in Foxp3⁺ cells only in the thymus. Levels of differentially expressed Foxp3 target genes in T_R , T_{FN} and 'activated' T_{25} cells (CD25⁺Foxp3⁻) were compared to naive CD25⁻Foxp3⁻ T_N cells⁴.

b, Foxp3-bound genes differentially expressed in T_R versus T_{FN} cells in both the thymus and periphery. Transcriptional profiling of indicated cell populations was performed using Affymetrix mouse 430 2.0 microarrays, and is described in detail elsewhere⁴.

target genes in *ex vivo* isolated T_R cells. As shown in Fig. 3c, permissive H3 histone modifications (acetylated H3K9/14 and trimethylated H3K4) are prevalent while inhibitory histone H3 modifications (tri-methylated H3K27, di- and tri-methylated H3K9) were not substantially changed in the corresponding regions of Foxp3 upregulated genes. On the other hand, the Foxp3 binding region of the repressed gene *Pde3b* was enriched for an inhibitory H3 modification (H3K27 tri-methylation) (Fig. 3c; data not shown). Thus, transcriptional activation and repression mediated by Foxp3 is associated with specific histone modifications at its binding sites.

Further analysis of the Foxp3-bound genes differentially expressed in the thymus revealed a group of genes encoding factors that control gene expression and chromatin modifications (34 of a total of 216 genes) as well as regulators of membrane and intracellular signalling (27 of a total of 216 genes) (Fig. 4a; Supplementary Table 3). This observation is consistent with the idea that these Foxp3 targets are involved in establishing a Foxp3-dependent differentiation programme. In contrast, examination of Foxp3-bound genes differentially expressed only in the periphery did not reveal obvious trends, perhaps owing to the relatively small size of this group. However, analysis of 35 genes that are subject to Foxp3-dependent modulation in both the thymus and periphery revealed a group of genes defining a T_R surface phenotype (*Il2ra*, *Ctla4*, *Nrp1*, *Icos*) and a group of genes encoding intracellular signalling regulators such as Map3k8 and PDE3b (Fig. 4b). Recently, we found an essential role for Foxp3-mediated *Pde3b* repression in the maintenance of T_R cells in the periphery by adapting them to chronic stimulation through TCR and IL-2 receptors⁴.

Notably, Foxp3 target genes comprise only a small portion (~6%) of the entire programme of Foxp3-dependent gene expression, suggesting that Foxp3 regulates a substantial part of the Foxp3-dependent transcriptional programme indirectly through other transcription factors. We found four transcription factor-encoding genes (*Prdm1*, *Crem*, *Zfpn1a2*, *Irf6*) among the 35 Foxp3-bound genes differentially expressed in both the thymus and periphery in addition to nuclear factor-encoding genes differentially expressed only in the thymus (*Stat5*, *Irf4*, *Hif1a*, *Stat4*) (Fig. 4a, b). These factors might be important for Foxp3-dependent developmental and functional programmes. In this regard, secondary Foxp3 targets probably include a

number of Foxp3-dependent genes encoding putative effector molecules of T_R mediated suppression such as *Gzmb*, *Fgl2*, *Il10* and *Ebi3*, which were not found within the Foxp3 direct target gene set.

Besides genic regions, we found 24% of Foxp3-binding peaks in intergenic regions (Fig. 2). This observation suggested that in addition to transcriptional regulation of protein-coding genes, Foxp3 regulates expression of non-coding RNA. To explore this possibility, we compared the relative locations of Foxp3 binding peaks with known mouse microRNA (miRNA), and found one prominent intergenic Foxp3-binding region in close proximity to the sequence encoding miR-155, a miRNA associated with *c-myc* upregulation, pre-B cell proliferation and tumorigenesis^{11–13}. Using Foxp3 ChIP-qPCR, we confirmed Foxp3 binding to this region in the *miR-155* locus in *ex vivo* isolated Foxp3⁺ T_R cells (Fig. 5a, b). To explore whether Foxp3 binding results in differential expression of *miR-155*, we used a qPCR approach to compare miR-155 levels in FACS-purified GFP⁺ T_{FN} versus T_R cells isolated from Foxp3^{gfpko/wt} and Foxp3^{gfp/gfp} female mice, respectively. In addition, miR-155 levels were measured in 'naive' (CD25⁺ Foxp3⁺) T_N cells and 'activated' (CD25⁺ Foxp3⁺) T_{25} cells. We found an ~7–8-fold increase in miR-155 in T_R cells compared to T_{FN} or T_N cells, while T_{25} cells showed a twofold increase in miR-155 (Fig. 5c). In agreement with these results, an increase of miR155 levels was reported in Foxp3-transduced non- T_R cells¹⁴. These results suggest that *miR-155* is a direct Foxp3 target. Its upregulation in T_R cells probably contributes to increased Foxp3-dependent expression of cell cycle genes and superior proliferative potential of T_R cells *in vivo*^{4,6}.

In conclusion, Foxp3 binds to ~700 genes and an intergenically encoded miRNA, and plays a dual role as both a transcriptional activator and repressor. Furthermore, Foxp3 probably sets up a transcription factor network controlling the overall functional programme of T_R cells. In agreement with this idea, genes responsible for regulating transcription and establishing epigenetic modifications were the most abundant among Foxp3 target genes in the thymus. Foxp3-mediated regulation of a small group of genes common to thymic and peripheral T_R cells affects a number of plasma membrane receptors and signalling molecules, as well as several transcription factors, and miR-155 miRNA. The latter group of Foxp3 targets probably plays an essential role in T_R survival, their proliferative potential, and suppressive function.

METHODS

Mice. C57BL/6 (B6) (the Jackson Laboratory), Foxp3 knockout (Foxp3^{-/-}), Foxp3-GFP protein reporter (Foxp3^{gfp})-expressing, and GFP-marked Foxp3 knockout allele (Foxp3^{gfpko})-expressing mice (B6 backcrosses 13, 9 and 5, respectively)^{4,6,15} were maintained in the University of Washington SPF animal facility.

Cell isolation. CD25⁺CD4⁺ T_R cells were purified from B6 lymph nodes and spleens at 4 °C using magnetic bead sorting as described elsewhere¹⁶. For histone modification ChIP, the CD4⁺CD25⁺ T cells were further purified using anti-CD4 Dynabeads (Dyna) to >98% purity. For experiments testing Foxp3 binding to the *miR-155* locus and PCR-based quantification of *miR-155* expression, T_R , T_{FN} , T_{25} and naive T_N cells were purified (>95% purity) from pooled lymph nodes and spleens of Foxp3^{gfp/gfp} and Foxp3^{gfpko/wt} female mice using a FACSaria sorter (BD Biosciences).

Chromatin immunoprecipitation. For Foxp3 ChIP, polyclonal Foxp3 antibodies were isolated from anti-Foxp3 rabbit IgG using Sepharose 4B-bound purified recombinant Foxp3 protein. 'Flow-through' IgG devoid of Foxp3 antibodies served as a negative control. Antibodies specific for different histone modifications were from Upstate Biotechnology. ChIP was carried out as described elsewhere¹⁷. Relative abundance of regions of interest in precipitated DNA was measured by qPCR using Power SYBR Green PCR master mix (Applied Biosystems). Primer sequences are listed in Supplementary Table 4.

Mouse genome tiling array analysis. Foxp3-bound DNA from purified B6 T_R and Foxp3^{-/-} CD4⁺ T cells (negative control) was PCR amplified, fragmented, labelled with the Affymetrix GeneChip WT Double-Stranded DNA Terminal Labelling Kit, and hybridized to the GeneChip Mouse Tiling 2.0R Array Set (Affymetrix). Raw array data were processed using the MAT software package to calculate peak intensity and location¹⁸. Upon comparison of the two samples, Foxp3-bound genes were identified by the location of each peak with *P* value <10⁻⁶ within 50 kb upstream or downstream of the predicted transcription start

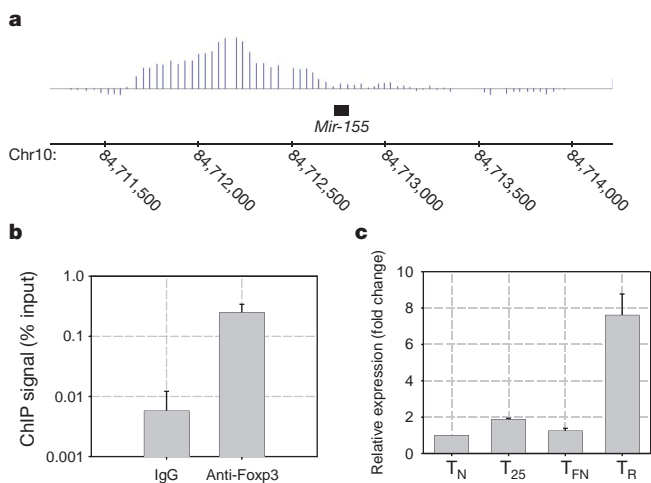


Figure 5 | Foxp3 regulates *miR-155* expression in T_R cells. **a**, Foxp3 binding ~500 bp upstream of the *miR-155* coding sequence was visualized using the Affymetrix Integrated Genome Browser. **b**, Confirmation of Foxp3 binding by ChIP coupled with qPCR. **c**, Increased *miR-155* expression in T_R cells. T_N (CD25⁺ Foxp3⁺), T_{25} (CD25⁺ Foxp3⁺) and T_R cells were FACS sorted from female Foxp3^{gfp/gfp} mice. T_{FN} cells were sorted from female Foxp3^{gfpko/wt} mice. Expression of *miR-155* was detected by a modified qPCR procedure (see Methods). All samples were compared to T_N cells to calculate relative miR-155 levels. Results are representative of two independent experiments; error bars, ± 1 s.d.

site of the closest gene. If transcription start sites of more than one gene were within these 50-kb intervals, the gene closest to the Foxp3 binding peak was taken into account.

Quantitative RT-PCR for miRNA. Total RNA was extracted using Stat-60 (IsoTex Diagnostics). First strand complementary DNA was synthesized by using the NCode miRNA First-Strand cDNA Kit (Invitrogen). Levels of miRNA were measured by qPCR using a miRNA specific forward primer and a universal reverse primer. Ubiquitously expressed U6 small nuclear RNA was used for normalization.

Received 30 October; accepted 21 December 2006.

Published online 21 January 2007.

- Fontenot, J. D. & Rudensky, A. Y. A well adapted regulatory contrivance: regulatory T cell development and the forkhead family transcription factor Foxp3. *Nature Immunol.* **6**, 331–337 (2005).
- Sakaguchi, S. *et al.* Foxp3⁺ CD25⁺ CD4⁺ natural regulatory T cells in dominant self-tolerance and autoimmune disease. *Immunol. Rev.* **212**, 8–27 (2006).
- Ziegler, S. F. FOXP3: of mice and men. *Annu. Rev. Immunol.* **24**, 209–226 (2006).
- Gavin, M. A. *et al.* Foxp3-dependent programme of regulatory T-cell differentiation. *Nature* doi:10.1038/nature05543 (in the press).
- Couronne, O. *et al.* Strategies and tools for whole-genome alignments. *Genome Res.* **13**, 73–80 (2003).
- Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341 (2005).
- Zhang, X. *et al.* Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc. Natl Acad. Sci. USA* **102**, 4459–4464 (2005).
- Beima, K. M. *et al.* T-bet binding to newly identified target gene promoters is cell type-independent but results in variable context-dependent functional effects. *J. Biol. Chem.* **281**, 11992–12000 (2006).
- Wu, Y. *et al.* FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell* **126**, 375–387 (2006).
- Chen, C., Rowell, E. A., Thomas, R. M., Hancock, W. W. & Wells, A. D. Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. *J. Biol. Chem.* **281**, 36828–36834 (2006).
- Tam, W., Hughes, S. H., Hayward, W. S. & Besmer, P. Avian bic, a gene isolated from a common retroviral site in avian leukosis virus-induced lymphomas that encodes a noncoding RNA, cooperates with *c-myc* in lymphomagenesis and erythroleukemogenesis. *J. Virol.* **76**, 4275–4286 (2002).
- Eis, P. S. *et al.* Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc. Natl Acad. Sci. USA* **102**, 3627–3632 (2005).
- Costinean, S. *et al.* Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in Eμ-miR155 transgenic mice. *Proc. Natl Acad. Sci. USA* **103**, 7024–7029 (2006).
- Cobb, B. S. *et al.* A role for Dicer in immune regulation. *J. Exp. Med.* **203**, 2519–2527 (2006).
- Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4⁺CD25⁺ regulatory T cells. *Nature Immunol.* **4**, 330–336 (2003).
- Gavin, M. A., Clarke, S. R., Negrou, E., Gallegos, A. & Rudensky, A. Homeostasis and anergy of CD4⁺CD25⁺ suppressor T cells *in vivo*. *Nature Immunol.* **3**, 33–41 (2002).
- Michael, L. F., Asahara, H., Shulman, A. I., Kraus, W. L. & Montminy, M. The phosphorylation status of a cyclic AMP-responsive activator is modulated via a chromatin-dependent mechanism. *Mol. Cell. Biol.* **20**, 1596–1603 (2000).
- Johnson, W. E. *et al.* Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl Acad. Sci. USA* **103**, 12457–12462 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Forbush and L. Karpik for assistance, and the members of our laboratory for discussions. This work was supported by the Cancer Research Institute (Y.Z.) and grants from the National Institutes of Health (A.Y.R.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.Y.R. (aruden@u.washington.edu).

TRAPPI tethers COPII vesicles by binding the coat subunit Sec23

Huaqing Cai^{1,2}, Sidney Yu^{1,2}, Shekar Menon^{1,2}, Yiyi Cai², Darina Lazarova¹, Chunmei Fu², Karin Reinisch², Jesse C. Hay³ & Susan Ferro-Novick^{1,2}

The budding of endoplasmic reticulum (ER)-derived vesicles is dependent on the COPII coat complex¹. Coat assembly is initiated when Sar1-GTP recruits the cargo adaptor complex, Sec23/Sec24, by binding to its GTPase-activating protein (GAP) Sec23 (ref. 2). This leads to the capture of transmembrane cargo by Sec24 (refs 3, 4) before the coat is polymerized by the Sec13/Sec31 complex⁵. The initial interaction of a vesicle with its target membrane is mediated by tethers⁶. We report here that in yeast and mammalian cells the tethering complex TRAPPI (ref. 7) binds to the coat subunit Sec23. This event requires the Bet3 subunit. *In vitro* studies demonstrate that the interaction between Sec23 and Bet3 targets TRAPPI to COPII vesicles to mediate vesicle tethering. We propose that the binding of TRAPPI to Sec23 marks a coated vesicle for fusion with another COPII vesicle or the Golgi apparatus. An implication of these findings is that the intracellular destination of a transport vesicle may be determined in part by its coat and its associated cargo.

Vesicle tethers work in conjunction with Rab GTPases⁶. In the yeast *Saccharomyces cerevisiae*, COPII vesicle tethering requires TRAPPI, the Rab Ypt1p (the suffix 'p' indicates a yeast protein) and a long coiled-coil protein called Uso1p (refs 7, 8). TRAPPI binds to COPII vesicles *in vitro* in the absence of Ypt1p and Uso1p (ref. 7). However, the subunit of TRAPPI that binds to these vesicles and the TRAPPI-binding partner on vesicles are unknown.

The TRAPPI subunit, Bet3p, is 54% identical to its mammalian counterpart (mBet3)⁹. mBet3 is recruited to the site where COPII vesicles bud to mediate the tethering of COPII vesicles to each other (homotypic tethering)¹⁰. Immuno-electron-microscopy demonstrating that mBet3 and the coat subunit mSec31 reside on the same vesicle¹⁰, prompted us to determine if mBet3 binds directly to COPII. We incubated increasing concentrations of His₆-mBet3 with glutathione *S*-transferase (GST)-tagged COPII subunits (0.5 μ M) and found that His₆-mBet3 bound specifically to beads containing GST-mSec23 (Fig. 1a). Saturable binding was achieved between concentrations of 3–10 μ M His₆-mBet3, depending on the purified protein preparation (data not shown). No binding was observed to GST, GST-mSec13, GST-mSec24, GST-CC1 (specificity control of the dynein-binding fragment of p150^{GLUED})¹¹ or the carboxy terminus of mSec31 (C-terminal 141 amino acids, data not shown). Furthermore, no specific binding was observed with other TRAPPI subunits (mTrs20, mBet5 and mTrs31, data not shown). We also demonstrated that mSec23 and mBet3 interact with each other *in vivo* (Supplementary Fig. 1a).

We determined if mSec23 has a role in tethering using the homotypic COPII tethering assay¹². COPII vesicle populations, marked in two different ways, were formed from permeabilized NRK cells. One population of vesicles contained Myc-tagged vesicular stomatitis

virus glycoprotein ts045 (VSV-G-Myc) and the other contained untagged, radiolabelled VSV-G (VSV-G*). The two vesicle populations were mixed and incubated together with cytosol and an ATP-regenerating system. VSV-G*-containing vesicles that tethered to VSV-G-Myc-containing vesicles were immunoprecipitated with anti-Myc antibody and detected by autoradiography. We found that excess soluble GST-mSec23 (Fig. 1b), but neither GST (Fig. 1b and 1c) nor GST-mSec24 (Fig. 1c), blocked tethering. Furthermore, both mSec31 (Fig. 1b, inset) and mSec23 (Fig. 1d) were detected on the co-isolated vesicles. We also compared the signal of mSec23 to VSV-G on vesicles formed in the presence or absence of GMP-PNP. GMP-PNP is a non-hydrolysable analogue of GTP that locks the coat on vesicles by preventing the hydrolysis of GTP on Sar1. COPII vesicles formed in the presence of GMP-PNP are maximally coated¹. Consistent with a functional role for mSec23 in homotypic COPII tethering, we estimated that 58% of the mSec23 was retained on the co-isolated vesicles when the cargo VSV-G was used to normalize to vesicle yield (Fig. 1d, lanes 2 and 3; Supplementary Information).

To address directly if TRAPPI binds to ER-derived vesicles via an interaction with Sec23, we initiated studies in yeast in which an assay that measures the binding of TRAPPI to COPII vesicles is available⁷. Consistent with the possibility that yeast TRAPPI (yTRAPPI) binds to Sec23p, we found that the *bet3-2* mutation causes synthetic growth defects in combination with the *sec23-1* mutation, but not with the *sec13-1* mutation (data not shown). *In vitro*, yTRAPP bound to GST-Sec23p, but not to GST, GST-Sec13p or GST-Ypt51p (Supplementary Fig. 1b). To assess the nature of this interaction, COPII vesicles, containing radiolabelled pro- α -factor, were formed *in vitro* from yeast donor membranes and incubated with Protein-A-tagged yTRAPPI that was immobilized on IgG sepharose beads. When increasing concentrations of GST-Sec23p, GST, GST-Sec13p or the C-terminus of GST-Gea2p (amino acids 972–1459 of the Arf1p exchange factor¹³) were added to the binding reaction, only GST-Sec23p inhibited vesicle binding in a concentration-dependent fashion (Fig. 2a). COPII vesicle binding to yTRAPPI was also inhibited when excess Bet3p-His₆, but not excess TRAPP subunit (Trs) Trs31p tagged with maltose binding protein (MBP) was included in the assay (Fig. 2b). This finding is consistent with the observation that mBet3, but not mTrs31, binds to GST-mSec23. In support of the proposal that yTRAPPI binds to Sec23p on vesicles, excess GST-Sec23p also inhibited the binding of yTRAPPI to COPII vesicles in the absence of cytosol (Fig. 2c; Supplementary Information).

To demonstrate directly that Sec23p is the binding partner for yTRAPPI on COPII vesicles, we used the temperature-sensitive *sec23-1* mutant to inactivate Sec23p on vesicles before performing the binding assay. Vesicles formed at 17 °C from *sec23* mutant

¹Howard Hughes Medical Institute and the ²Department of Cell Biology Yale University School of Medicine, New Haven, Connecticut 06519, USA. ³Division of Biological Sciences, The University of Montana, Missoula, Montana 59812-4824, USA.

fractions showed a temperature-sensitive defect in binding to γ TRAPPI (Fig. 2d, compare binding at 17 °C and 29 °C). Together our findings demonstrate that γ TRAPPI binds to ER-derived vesicles through an interaction between Sec23p and Bet3p. Neither γ TRAPPI nor recombinant Bet3p affected the GAP activity of Sec23p/Sec24p when measured with the tryptophan fluorescence assay¹⁴ (data not shown). This and other data (see Supplementary Information) indicate that the GAP activity of Sec23p is not required for the binding of γ TRAPPI to COPII vesicles (Supplementary Fig. 2).

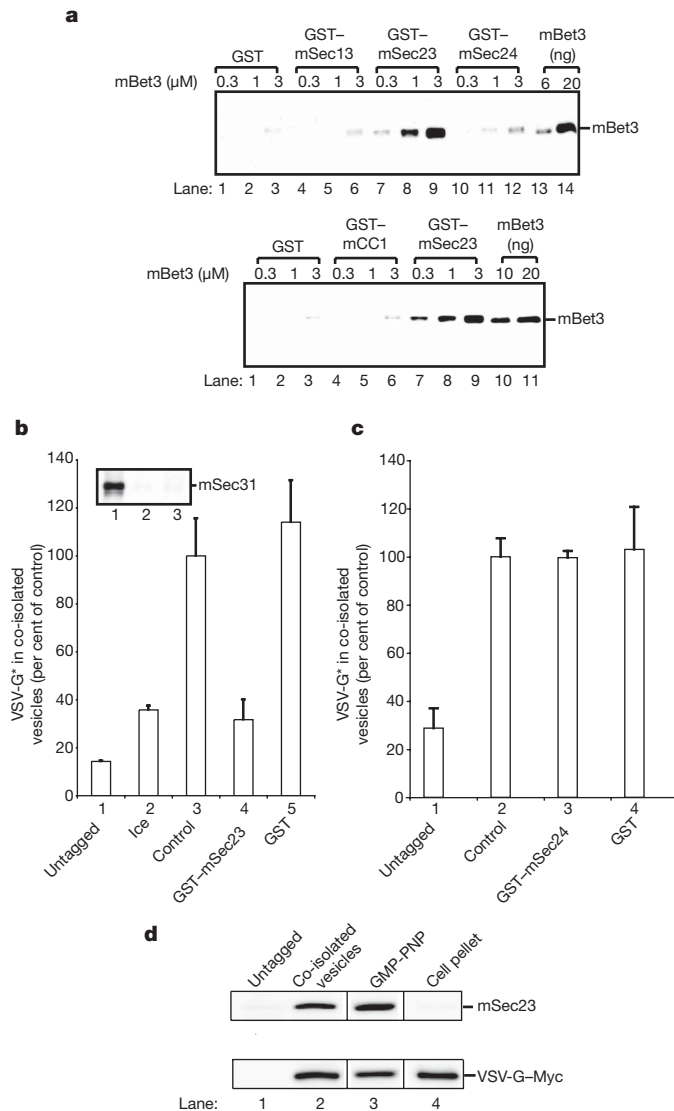


Figure 1 | mBet3 tethers coated vesicles in mammalian cells. **a**, mBet3 binds to mSec23 *in vitro*. mBet3 at 6 and 20 ng (lanes 13 and 14, top panel) and 10 and 20 ng (lanes 10 and 11, bottom panel) were used as standards. **b**, mSec23 mediates homotypic COPII vesicle tethering. Two COPII vesicle populations, one marked with VSV-G-Myc and another with radiolabelled VSV-G (VSV-G*), were incubated together in the absence (lanes 2 and 3) or presence of 0.4 μ M GST-mSec23 (lane 4) or 0.4 μ M GST (lane 5) at 32 °C, or on ice (lane 2). One sample contained untagged vesicles and VSV-G*-containing vesicles (lane 1). mSec31 is present on nascent vesicular tubular clusters formed *in vitro* (inset; lane 1, complete reaction; lane 2, no vesicle control (see Supplementary Information); lane 3, untagged control (see Supplementary Information)). **c**, The experiment was performed as in **b** with 0.4 μ M GST (lane 4) and 0.4 μ M GST-mSec24 (lane 3). **d**, mSec23 is present on co-isolated vesicles. Lane 1, untagged control; lane 2, complete reaction; lane 3, vesicles produced and isolated in the presence of 100 μ M GMP-PNP; lane 4, 0.67% of input cells used to generate material in lanes 2 and 3. Plotted are the means of duplicate determinations \pm s.e.m. Similar results were obtained in two independent experiments.

To determine if the interaction between Sec23p and γ TRAPPI is required for vesicle tethering in yeast, we performed the *in vitro* ER-to-Golgi tethering assay¹⁵. Vesicles formed in the first stage of this assay were incubated in a second stage with Golgi membranes, cytosol and an ATP-regenerating system. Whereas COPII vesicle

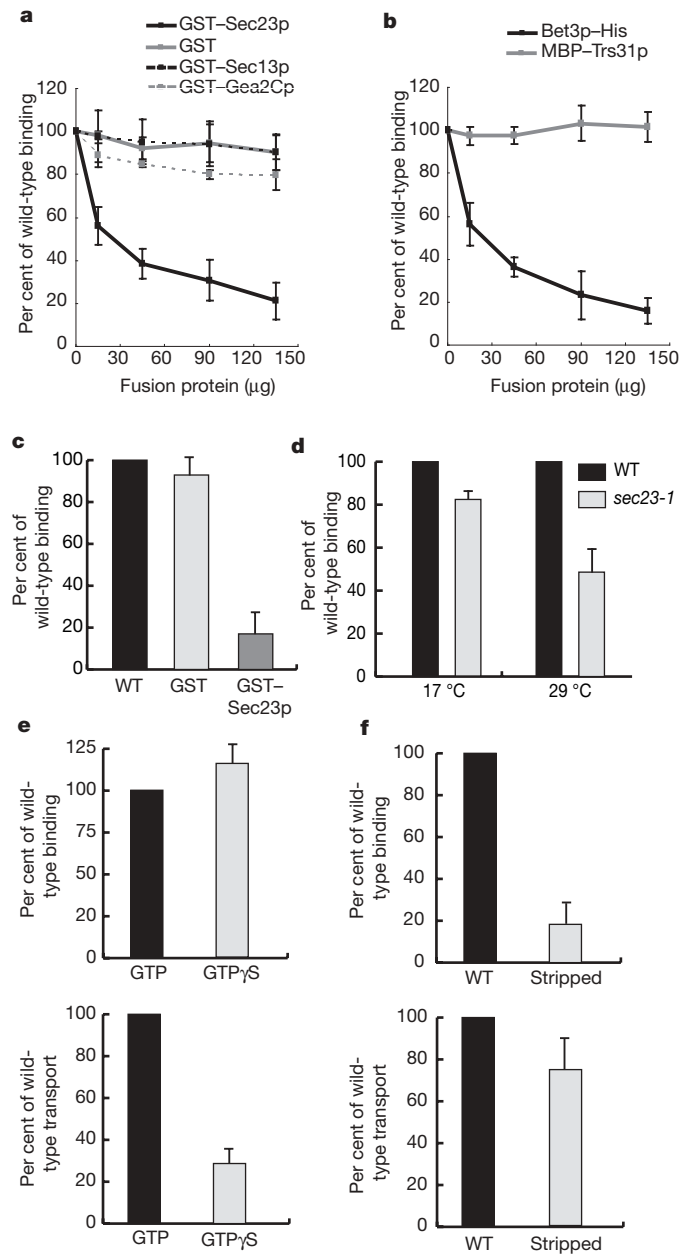


Figure 2 | The interaction between Sec23p and Bet3p is required for the binding of COPII vesicles to γ TRAPPI. **a**, GST-Sec23p, but not GST, GST-Sec13p or GST-Gea2Cp, inhibits the binding of COPII vesicles to γ TRAPPI. **b**, Bet3p-His₆, but not MBP-Trs31p, inhibits the binding of COPII vesicles to γ TRAPPI. **c**, GST-Sec23p, but not GST, blocked the binding of γ TRAPPI to cytosol-free COPII vesicles prepared as described in Supplementary Information. **d**, Vesicles prepared from *sec23-1* mutant cells exhibit a temperature-sensitive defect in vesicle binding. **e**, Top panel, vesicles formed in the presence of GTP γ S bind efficiently to γ TRAPPI. Bottom panel, the concentration of GTP γ S (100 μ M) used to generate vesicles efficiently blocks transport. **f**, Vesicles stripped of COPII coats, as described in Supplementary Information, fail to bind to γ TRAPPI (top panel), even though they are fusion-competent in the presence of cytosol (bottom panel). Approximately equal amounts of vesicles were used in all bindings. Error bars are s.d. In all experiments $n = 2$, except for experiments in **a** and **b** with GST-Sec23p and Bet3p-His₆ in which $n = 4$.

tethering is largely a homotypic event in mammalian cells, in yeast, COPII vesicles are thought to tether and fuse directly with the Golgi (heterotypic fusion)^{12,15,16}. Subsequent to the second stage incubation, the reaction product was fractionated on a sucrose velocity gradient and the migration of vesicles was monitored with ConA sepharose, which binds to glycosylated vesicle cargo (pro- α -factor). Free vesicles peaked in fraction 5 and were separated from tethered vesicles (fraction 8; Fig. 3a). When fusion occurred, pro- α -factor was modified with outer chain carbohydrate in the Golgi (fraction 8; Fig. 3b). Vesicle tethering and fusion were unaffected in the presence of excess soluble GST or GST–Sec13p (fraction 8; Fig. 3a, b). Similarly, GST–Sec24p did not block vesicle tethering (Supplementary Fig. 3a). In contrast, both tethering and fusion were completely blocked in the presence of excess soluble GST–Sec23p (Fig. 3a, b). Additionally, as predicted from our vesicle binding experiments (Fig. 2d), the *sec23-1* mutant showed a temperature-sensitive defect in vesicle tethering (Supplementary Fig. 3b and 3c). Together these findings indicate that γ TRAPPI tethers COPII vesicles via an interaction with Sec23p.

Our data imply that vesicle tethering precedes vesicle uncoating. *In vitro* the non-hydrolysable GTP analogue, GTP γ S, blocks ER–Golgi traffic (Fig. 2e, bottom panel) by preventing the hydrolysis of GTP on Sar1, which locks the coat on vesicles (Supplementary Fig. 4a). Sar1 dissociates from membranes when GTP is hydrolysed^{1,17}. However, the ability to bind cargo allows the Sec23/Sec24 complex to remain transiently associated with vesicles after hydrolysis^{1,18,19}. We found that γ TRAPPI bound efficiently to vesicles formed in the presence of GTP γ S (Fig. 2e, top panel). Thus, the binding of γ TRAPPI to ER-derived vesicles is not inhibited by blocking the hydrolysis of GTP on Sar1p. This observation is consistent with the proposal that the GAP activity of Sec23p is not required for COPII vesicle tethering. Furthermore, when we stripped vesicles of their coat by sedimentation through a sorbitol cushion (Supplementary Fig. 4b), γ TRAPPI failed to bind to them (Fig. 2f, top). Stripped vesicles were still fusion competent (Fig. 2f, bottom) when assayed in the presence of cytosol, which

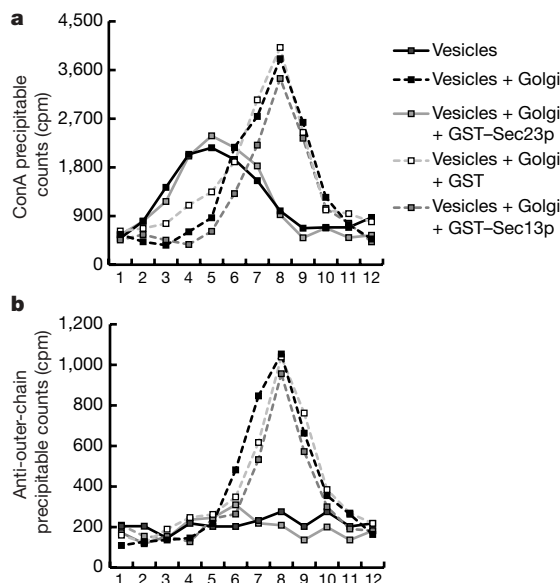


Figure 3 | Sec23p is required for COPII vesicle tethering. Tethered vesicles peaked in fraction 8 (filled box, broken line in **a**) with Golgi membranes that contained outer chain carbohydrate (**b**). cpm, counts per minute. **a**, Excess GST–Sec23p blocked vesicle tethering (grey box, solid line) and membrane-bound pro- α -factor migrated on a sucrose velocity gradient with free vesicles that peaked in fraction 5 (black box, solid line). The conditions used for the gradient have been described before¹⁵. **b**, In the presence of GST–Sec23p, the pro- α -factor within vesicles was not modified with outer chain carbohydrate. Excess GST (open box, broken line) or GST–Sec13p (grey box, broken line) did not disrupt vesicle tethering (**a**) or fusion (**b**).

may allow coat proteins to be recruited back onto them. Together these findings demonstrate that the binding of γ TRAPPI to ER-derived vesicles is dependent on the presence of the COPII coat.

Although it is known that γ TRAPPI activates Ypt1p by converting it to its GTP-bound form, the subunit(s) of the complex required for this event is unknown⁷. If the interaction between Bet3p and Sec23p is sufficient for γ TRAPPI-dependent COPII vesicle tethering, then Bet3p may mediate nucleotide exchange on Ypt1p. However, when we tested the ability of Bet3p–His₆ to stimulate nucleotide exchange on Ypt1p, no activity was detected (Fig. 4b). To identify the minimum number of subunits that are needed to reconstitute nucleotide exchange, we co-expressed Bet3p with other γ TRAPPI subunits in *Escherichia coli*, purified the resulting complexes and analysed them for exchange activity (Fig. 4). The analysis of the

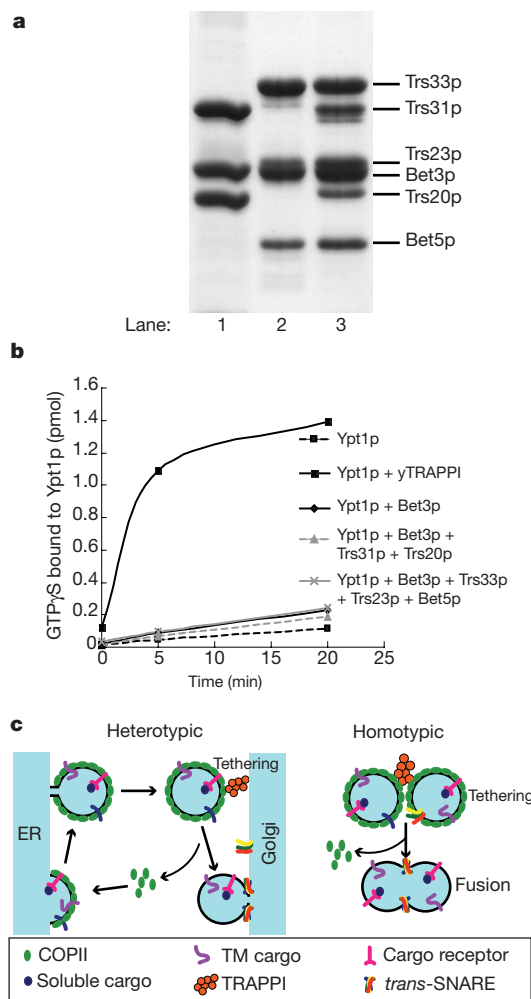


Figure 4 | γ TRAPPI links vesicle recognition to Rab activation. **a**, γ TRAPPI subunits were expressed in bacteria to generate three different Bet3p-containing complexes that were purified and resolved on a 14% SDS polyacrylamide gel and stained with Coomassie blue. Lane 1, Trs31p, Bet3p and Trs20p; lane 2, Trs33p, Trs23p, Bet3p and Bet5p; lane 3, Trs33p, Trs31p, Trs23p, Bet3p, Trs20p and Bet5p. **b**, GTP γ S uptake onto Ypt1p was measured as described before²¹. In the presence of 40 nM γ TRAPPI and 160 nM Ypt1p, a 9.6-fold stimulation in the rate of uptake of GTP γ S was observed. All assays were performed with equivalent amounts of Bet3p. **c**, A model for heterotypic and homotypic COPII vesicle tethering events. In heterotypic tethering, γ TRAPPI tethers ER-derived vesicles to the Golgi through a direct interaction with Sec23p. Coat complexes dissociate from vesicles, followed by the formation of trans-SNARE pairs and membrane fusion. In homotypic vesicle tethering, TRAPPI links two COPII-coated vesicles via an interaction between mBet3 and mSec23. This model is consistent with our earlier observation²⁸ that there is more than one copy of Bet3p in the complex (also see note in proof). TM, transmembrane.

subunit interactions of Bet3, with the low molecular weight subunits of TRAPPI²⁰, enabled us to divide yTRAPPI into two Bet3p-containing subcomplexes. One subcomplex contained Bet3p, Trs20p and Trs31p, whereas the other included Bet3p with Trs33p, Trs23p and Bet5p (Fig. 4a). Neither subcomplex had exchange activity (Fig. 4b). However, when all six yTRAPPI subunits were expressed in bacteria, they assembled into a complex that stimulated nucleotide exchange on Ypt1p (Fig. 4b). Complex assembly and exchange activity were not dependent on Trs33p (data not shown), which is dispensable for growth in yeast⁹. The other non-essential TRAPPI subunit, Trs85p, is not required for Ypt1p exchange activity in yeast²¹. Thus, all five essential yTRAPPI subunits are required to coordinate the initial interaction of the vesicle with the activation of Ypt1p, an event that is needed to stably tether the vesicle to its target membrane.

Until now, the recognition of a vesicle by the tethering/fusion machinery was thought to occur after the vesicle uncoats²². It has been postulated that the fusion selectivity of a vesicle with its target membrane requires the SNAREs, a highly conserved family of membrane proteins, which are exposed when the vesicle uncoats²³. Cognate SNAREs, residing on the vesicle and target membrane, must pair for fusion to take place²⁴. Surprisingly, we have shown that a component, Sec23, of the COPII cargo adaptor complex links cargo sorting to target recognition and tethering. These findings have revealed an additional role for vesicle coats. Our data indicates that ER-derived vesicles recognize and bind to their target membrane before uncoating (Fig. 4c). Once the vesicle uncoats, *trans*-SNARE pairing and membrane fusion can occur²⁴. Uso1p, like its mammalian homologue p115, may mediate the tethering of uncoated vesicles via an interaction with the SNAREs^{25,26}.

Note added in proof: The crystal structure of TRAPPI was reported during the revision of this manuscript²⁷. TRAPPI contains two lobes and a long positively charged flat surface. Each lobe of the complex contains one copy of Bet3. This structure is consistent with the homotypic tethering model proposed in Fig. 4c, where each copy of Bet3 interacts with a different coated vesicle.

METHODS

In vitro binding experiment with mBet3 and COPII subunits. The expression of all recombinant mammalian proteins and the conditions used for the *in vitro* binding assay are described in Supplementary Information.

Homotypic tethering assay. The mammalian homotypic COPII vesicle tethering assay was performed as described before¹². More detail is provided in Supplementary Information.

COPII vesicle binding studies and yeast vesicle tethering assay. Vesicle binding experiments with yeast COPII vesicles and the yeast tethering assay were performed as described before^{7,15}. More detail is provided in Supplementary Information.

Received 4 October; accepted 11 December 2006.

Published online 7 February 2007.

- Barlowe, C. *et al.* COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* **77**, 895–907 (1994).
- Yoshihisa, T., Barlowe, C. & Schekman, R. Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum. *Science* **259**, 1466–1468 (1993).
- Kuehn, M. J., Herrmann, J. M. & Schekman, R. COPII-cargo interactions direct protein sorting into ER-derived transport vesicles. *Nature* **391**, 187–190 (1998).
- Miller, E. A. *et al.* Multiple cargo binding sites on the COPII subunit Sec24p ensure capture of diverse membrane proteins into transport vesicles. *Cell* **114**, 497–509 (2003).
- Stagg, S. M. *et al.* Structure of the Sec13/31 COPII coat cage. *Nature* **439**, 234–238 (2006).

- Whyte, J. R. & Munro, S. Vesicle tethering complexes in membrane traffic. *J. Cell Sci.* **115**, 2627–2637 (2002).
- Sacher, M. *et al.* TRAPP I implicated in the specificity of tethering in ER-to-Golgi transport. *Mol. Cell* **7**, 433–442 (2001).
- Cao, X., Ballew, N. & Barlowe, C. Initial docking of ER-derived vesicles requires Uso1p and Ypt1p but is independent of SNARE proteins. *EMBO J.* **17**, 2156–2165 (1998).
- Sacher, M. *et al.* TRAPP, a highly conserved novel complex on the cis-Golgi that mediates vesicle docking and fusion. *EMBO J.* **17**, 2494–2503 (1998).
- Yu, D. *et al.* mBet3p is required for homotypic COPII vesicle tethering in mammalian cells. *J. Cell Biol.* **174**, 359–368 (2006).
- Quintyne, N. J. *et al.* Dynactin is required for microtubule anchoring at centrosomes. *J. Cell Biol.* **147**, 321–334 (1999).
- Xu, D. & Hay, J. C. Reconstitution of COPII vesicle fusion to generate a pre-Golgi intermediate compartment. *J. Cell Biol.* **167**, 997–1003 (2004).
- Peyroche, A. & Jackson, C. L. Functional analysis of ADP-ribosylation factor (ARF) guanine nucleotide exchange factors Gea1p and Gea2p in yeast. *Methods Enzymol.* **329**, 290–300 (2001).
- Antonny, B., Madden, D., Hamamoto, S., Orci, L. & Schekman, R. Dynamics of the COPII coat with GTP and stable analogues. *Nature Cell Biol.* **3**, 531–537 (2001).
- Barrowman, J., Sacher, M. & Ferro-Novick, S. TRAPP stably associates with the Golgi and is required for vesicle docking. *EMBO J.* **19**, 862–869 (2000).
- Rexach, M. F. & Schekman, R. W. Distinct biochemical requirements for the budding, targeting, and fusion of ER-derived transport vesicles. *J. Cell Biol.* **114**, 219–229 (1991).
- Oka, T. & Nakano, A. Inhibition of GTP hydrolysis by Sar1p causes accumulation of vesicles that are a functional intermediate of the ER-to-Golgi transport in yeast. *J. Cell Biol.* **124**, 425–434 (1994).
- Sato, K. & Nakano, A. Dissection of COPII subunit-cargo assembly and disassembly kinetics during Sar1p-GTP hydrolysis. *Nature Struct. Mol. Biol.* **12**, 167–174 (2005).
- Forster, R. *et al.* Secretory cargo regulates the turnover of COPII subunits at single ER exit sites. *Curr. Biol.* **16**, 173–179 (2006).
- Menon, S. *et al.* mBet3 is required for the organization of the TRAPP complexes. *Biochem. Biophys. Res. Commun.* **350**, 669–677 (2006).
- Wang, W., Sacher, M. & Ferro-Novick, S. TRAPP stimulates guanine nucleotide exchange on Ypt1p. *J. Cell Biol.* **151**, 289–296 (2000).
- Bonifacio, J. S. & Glick, B. S. The mechanisms of vesicle budding and fusion. *Cell* **116**, 153–166 (2004).
- Söllner, T. *et al.* SNAP receptors implicated in vesicle targeting and fusion. *Nature* **362**, 318–324 (1993).
- Rothman, J. E. Mechanisms of intracellular protein transport. *Nature* **372**, 55–63 (1994).
- Shorter, J., Beard, M. B., Seemann, J., Dirac-Svestrup, A. B. & Warren, G. Sequential tethering of Golgins and catalysis of SNAREpin assembly by the vesicle-tethering protein p115. *J. Cell Biol.* **157**, 45–62 (2002).
- Bentley, M., Liang, Y., Mullen, K., Xu, D., Sztul, E. & Hay, J. C. SNARE status regulates tether recruitment and function in homotypic COPII vesicle fusion. *J. Biol. Chem.* **281**, 38825–38833 (2006).
- Kim, Y.-G. *et al.* The architecture of the multisubunit TRAPP I complex suggests a model for vesicle tethering. *Cell* **127**, 817–830 (2006).
- Sacher, M., Barrowman, J., Schieltz, D., Yates, J. R. III & Ferro-Novick, S. Identification and characterization of five new subunits of TRAPP. *Eur. J. Cell Biol.* **79**, 71–80 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Walker for help in the preparation of the manuscript, Y. Zhang, and K. Mullen for technical assistance. We also thank J. Barrowman for help with the yeast *in vitro* tethering assay, R. Schekman, C. Kaiser, F. Gorelick and T. Schroer for purified proteins, antibodies, plasmids and strains, and E. Miller for advice on stripping COPII coats from vesicles. This work was supported by the Howard Hughes Medical Institute. Salary support for S.F.-N., H.C., S.Y., S.M. and D.L. was also provided by the Howard Hughes Medical Institute. Work at the University of Montana was supported by an NIH grant to J.C.H. and the NIH-COBRE Center for Structural and Functional Neuroscience.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.F.-N. (susan.ferronovick@yale.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

The film *Lost in Translation* features two Americans in Tokyo whose alienation is heightened not just by language difficulties (neither speaks Japanese) but also by cultural differences. Both characters arrive in Japan feeling some form of disquiet. Bill Murray's character is an actor whose career is in decline and who has come to Tokyo to film some lucrative whiskey commercials. Scarlett Johansson's character has come to Japan with her photographer husband, who leaves her in the hotel while he goes to photo shoots.

The article 'Lost in Translation' (see *Nature* **445**, 454–455; 2007) also centred on the issue of language differences. It looked at the difficulties expatriates experience in labs where their adopted language dominates. It also emphasized some of the cultural differences such experiences cause, which can get in the way of scientific productivity, particularly publishing.

Responses to the article have seen a difference of opinion among scientists over who should be responsible for saving science from being 'lost in translation'. Santanu Dasgupta of the Department of Cell and Molecular Biology at Uppsala University in Sweden places some responsibility on journal editors and referees who "are not doing their jobs, either from laziness or parochial arrogance or self interest". Janet Carter-Sigglow, a translator at a German research centre, says that non-native speakers should draw on professional translators like herself. And Francesco Colucci, of the Babraham Institute in Cambridge, UK, puts responsibility for communication squarely on scientists themselves — but emphasizes that everyone could benefit from understanding cultural differences better, which can't always be translated as easily as linguistic differences.

To help scientists get past barriers to publishing, *Nature* has launched a web page for existing and potential authors (www.nature.com/authors/author_services/how_write.html) that we hope will help prevent good research from being lost in translation. This and other matters are up for discussion on Nautilus, our author blog, at <http://blogs.nature.com/nautilus>.

Paul Smaglik, *Naturejobs* editor

CONTACTS

Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
UK/Ireland/Italy/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

Advertising Production Manager:
Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development:
Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street,
9th Floor,
New York,
NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

CORBIS

2-23-03

Fusion Proteins

Size of DNA fragments

PCR #	Size of DNA	Primer #	Comment
# 1	843-370 = 467 bp	47, 48	HR N-to
# 2	644-105 = 539 bp	49, 50	HC12 N-to
# 3	2421-1350 = 1071 bp	51, 52	HC12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100
# 4 good	732-340 = 392 bp	53, 54	HC12 N-to
# 5 good	2068-1477 = 591 bp	55, 56	HC12 N-to
# 6	1972-1576 = 396 bp	57, 58	HR CN-12
# 7	2433-1546 = 887 bp	59, 60	HR C-12
# 8	2831-1710 = 1121 bp	61, 62	HC 12
# 9	2331-1477 = 854 bp	63, 55	HC12

EST = 27 bp

4 = 392 bp + 27 bp

5 = 591 bp + 27 bp

COUNT ON IT

It's a tough brief to fill, but geneticists with mathematical and computing skills find jobs easy to get. Virginia Gewin reports.

Andrew Skol is in an enviable position. Unlike many biology postdocs facing a possible lifetime of temporary jobs, Skol went straight to a faculty-level instructor position in the University of Chicago's Department of Medicine after completing his PhD. While others generate the gene-expression profiles, Skol mines the morass of data for biological meaning — a skill that's increasingly in high demand.

As the catalogue of genetic variations grows, so does the ability to identify genes that influence traits of medical interest. DNA microarrays enable biologists to explore the most subtle of human genetic variations, single-nucleotide polymorphisms (SNPs) — differences of only one base pair between the same DNA sequence in two humans. The decreasing cost and increased performance of microarrays and similar technology, which are now able to screen for up to 500,000 SNPs at once, mean that statisticians well versed in genetics are now in demand in both industry and academia.

The 2005 publication of the International HapMap Project — a library of human genetic variation —

heralded the dawn of studies that aim to link disease susceptibility to complex genetic patterns. "The demand for people able to relate genetic variation to disease is exploding," says Lisa Brooks, director of the genetic variation programme at the National Human Genome Research Institute in Bethesda, Maryland.

Data provided by projects such as HapMap mean that scientists interested in linking disease to genotype will soon be able to dissect out the multiple causes of susceptibility to complex diseases such as heart disease and cancer. "The analysis focus of large-scale association studies will soon move beyond simple phenotypes, such as simply the presence or absence of cancer, to more complex phenotypes," says Christoph Lange, assistant professor of biostatistics at the Harvard School of Public Health in Boston, Massachusetts.

The increasing complexity of genetic-variation data is driving the need for more sophisticated statistical skills, says Bruce Weir, chair of the biostatistics department at the University of Washington in Seattle. It's not only gene-expression data that have to be taken account of. It is now possible, and desirable, to incorporate proteins and metabolites into the data mix.

Career opportunities in statistical genetics are plentiful, but training is not. Even for a good mathematician, gaining experience in the special blend of statistics, genetics and computer programming required might prove the biggest hurdle to landing a job in this field.

Training opportunities

The top statistical genetics departments in the United States focus primarily on training PhDs and postdocs, but some offer MS or certificate programmes emphasizing statistical genetics. The University of Alabama at Birmingham, for example, offers a certificate in statistical genetics only to those completing a PhD, while Washington University in St Louis, Missouri, offers a master's-level genetic epidemiology certificate with an MS degree. Whether you have an MS or a PhD could make a difference to the sort of job you can expect (see 'Look before you leap').

Despite few formal statistical genetics teaching programmes, there are signs that demand is beginning to fuel new training opportunities. Genetics

LOOK BEFORE YOU LEAP

Even when you get offered a job, there can be pitfalls ahead. Would you be happy simply serving the statistical needs of other researchers? "If that's not what you want and you are being hired as the only statistician in the place, you are in trouble," says David Allison, head of the statistical genetics section at the University of Alabama's biostatistics department.

Companies hire statistical geneticists trained to either MS or PhD level, but MS-level entrants can end up doing routine data analysis for methods that are being created by the PhDs in the development department. "The people who can design experiments to analyse data and interpret results and apply the findings are essential," says Meg Ehm, director of genetic epidemiology and analysis for pharmacogenetics at GlaxoSmithKline, based in London.

She says the company has struggled to find candidates with the appropriate statistics and genetics skills — sometimes taking up to 12 months to fill positions that require computer-programming skills as well. **V.G.**

departments are recruiting specialist instructors, anticipating that student interest won't be far behind. "We're scrambling to hire faculty to educate students about these opportunities," Weir says.

Even where opportunities exist, however, students don't seem to have caught on to the demand for statistical skills. "Often, people find us by chance, and only then get excited by the field," says Wolfgang Huber, statistics group leader at the European Bioinformatics Institute in Cambridge, UK. He is constantly on the lookout for people with a robust mathematical and biological background and the flexibility to learn about unfamiliar fields. He worries that the limited training in current methods that many people get will not serve their long-term career goals. "The field is moving so fast that narrow training might be obsolete five years from now."

Networking pays

In the small community of statistical genetics, the informal network is key. Skol was one of its beneficiaries. He did his PhD with biostatistician Michael Boehnke at the Center for Statistical Genetics at Michigan State University. As with all his students, Boehnke sent an e-mail to colleagues to let them know he had a student ready to leave. "There are many more positions than are even advertised in journals," says Boehnke.

Informal networks are important in Europe as well. "Europe has yet to put in place a much-needed statistics epicentre," says Leena Peltonen, molecular geneticist at the University of Helsinki and the National Public Health Institute in Helsinki, Finland. A few groups offer training in statistical genetics. Françoise Clerget-Darpoux at the National Institute for Health and Medical Research in Paris, France, trains between three and eight master's-level students each year, and she also helps to teach short courses organized by the European Genetics Foundation (see 'Keeping ahead of the game').

In Asia, where demand and student interest are increasing, Japan is the predominant training ground. Institutes in Singapore, for example, rely on statistical geneticists trained in North America or Europe. This is because educational infrastructure has not embraced the team science approach needed for statistical genetics, says Edison Liu, executive-director of the Genome Institute of Singapore.

Genetics statisticians are also urgently needed in industry. Pharmaceutical companies are increasingly turning to pharmacogenomics — the linking of individual genotypes to disease susceptibility or responses to drugs — to salvage the struggling business of drug development (see *Nature* **436**, 746–747; 2005). To do this they need to sift through the roughly 10 million SNPs in the human population to find the few that can explain an individual's susceptibility to disease or good or bad responses to drugs. Some companies are investing heavily in pharmacogenomic approaches and often partner with biotechnology companies developing the complex statistical analyses to identify potential drug targets.

"All the easy things, the low-hanging fruit, have been found in drug development," says John Hooper, chief executive of Genizon Biosciences, a Quebec-based company specializing in gene mapping and genetic analysis. Genizon provides genomic analysis services to pharmaceutical giant Pfizer, and is looking to expand

KEEPING AHEAD OF THE GAME

For those already in the field, short courses are a good way to keep your skills up to date. The European Genetics Foundation holds courses on statistical topics and the UK-based Wellcome Trust runs week-long statistics courses combining lectures with hands-on data-analysis practicals.

The University of Washington in Seattle hosts two summer institute training programmes in statistical genetics — one in Seattle and one in Liège, Belgium. The genetic analysis workshop held every two years by the Southwest Foundation for Biomedical Research in San Antonio, Texas, is another way to gain access to the latest analysis techniques, and a good opportunity to network.

Of the 345 participants last year, 43% were graduate students and postdocs. "Anybody can come, but only if they've done something with the data sets distributed beforehand," says Jean MacCluer (pictured), statistical geneticist and co-organizer of the workshop.

Genetic Analysis Workshop

♦ www.gaworkshop.org

Summer Institute in Statistical



Genetics at the University of Washington

♦ www.biostat.washington.edu/sisg07

European counterpart in Liège, Belgium

♦ www.functionalgenomics.org.uk/sections/activites/2007/Georges/info.htm

Wellcome Trust short course

♦ www.wellcome.ac.uk/doc_wtx026851.html

European Genetics Foundation

♦ www.eurogene.org/index.php



Christoph Lange (top) and Bruce Weir both see a bright future for mathematically minded geneticists.

its team of around 30 statisticians and computer programmers who mine genomic data for drug targets. "Statistics is the engine of our business," says Hooper.

Companies are looking in particular for statistical geneticists with computer-programming skills. Entry-level bioinformaticians sometimes think they can analyse complex genetic data well enough, but it's not so simple, says Eric Schadt, senior scientific director for genetics at Rosetta Inpharmatics, in Seattle, Washington state, a subsidiary of Merck. "When you are working with 500,000 SNPs, intersecting with 40,000 traits in 2,000 patients — that leads to terabytes of result data," says Schadt, who is actively recruiting statistical geneticists for Rosetta. "Moving through this scale of data requires top-notch programming skills."

Industry or academia?

Rosetta Inpharmatics evolved from a bioinformatics company into drug discovery, with an emphasis on methods-driven statistical development. Its high-profile publication record helps recruit scientists who may be unsure about leaving academia. For despite the roughly 30% pay hike that the pharmaceutical industry offers, as well as access to real-world data, many biostatisticians, like Skol, want to stay in academia.

Until pharmacogenomics proves its value, that may be a smart move. But, no matter how you slice the data, biology is becoming more analytical. "The days of molecular biologists as empirical scientists are past," says Hooper. "You can't eyeball 1.5 terabytes of data."

Skol agrees. He's got job security for the next few years, and he knows he is positioned for a future that will demand statistical skills. "As the lines between population genetics, bioinformatics, computational biology and statistics continue to blur, they will all merge into a broader field," he says.

Virginia Gewin is a freelance writer in Portland, Oregon.

MOVERS

Colin Masters, executive director, Mental Health Research Institute of Victoria, Australia



1989–2006: Consultant, chief of neuropathology laboratory, and chair, senior scientists' council, Mental Health Research Institute of Victoria
1989–2006: Consultant, Royal Melbourne Hospital
1999–2005: Associate dean (research), faculty of medicine, dentistry and health sciences, University of Melbourne

While a medical school student at the University of Western Australia, Colin Masters became fascinated by brain diseases above all else. He focused specifically on 'slow virus diseases' of the brain, including those with a delayed clinical onset such as prion diseases.

Masters first looked at prions to study the deposition of protein aggregates called amyloids. Seeing a possible link, he turned that work into a project studying the nature of amyloid deposits in Alzheimer's patients. He says the decision to research neurodegenerative diseases was the first pivotal moment of his career. At the time, the field adhered to a more descriptive study of neuropathology. With his collaborators, Masters capitalized on the newly available power of protein purification, identification and sequencing. Together, they sequenced the amyloid protein and began a fruitful decade of discoveries.

Having championed the amyloid theory of Alzheimer's, which holds that the accumulation of toxic amyloid protein impairs cognition, Masters is most proud of sparking an explosion of interest in the field. The amyloid protein is one of the most highly studied proteins in the whole of cell biology, boasting more than 800 papers annually.

In 2000, after running a laboratory at the Mental Health Research Institute of Victoria (MHRI) in Melbourne since the late 1980s, Masters turned his focus to the environmental and genetic contributors to Alzheimer's. His shift coincided with industry collaborations geared towards developing therapies for the disease, and the co-founding of a small biotech company called Prana Biotechnology. With one potential therapy currently undergoing early phase II drug trials in Sweden, Masters has turned his attention back to research.

As executive director of the MHRI, Masters plans to use his Alzheimer's experience to address psychoses and mood disorders. The synergy between the MHRI and the nearby Howard Florey Institute at the University of Melbourne should help.

"Colin Masters coming to Melbourne gives us the opportunity to bring molecular imaging and genetic techniques to psychiatric disorders, such as schizophrenia and bipolar disorder," says Fred Mendelsohn, director of the Howard Florey Institute.

Says Masters: "Clearly understanding how the brain and mind work are the biggest challenges for the next 10–20 years — and I want to be there as it is unfolding." ■
Virginia Gewin

NETWORKS & SUPPORT

Torn between two ladders

Much has been said about the pyramidal structure in academic science. But what if budding scientists discover halfway into their PhD that they're climbing the wrong career ladder? I was fortunate enough to find a way to change the focus of my career without having to compromise.

I believe that established scientists generally share a passion for science. Like Archimedes making his most famous discovery while bathing, their research questions stay with them even outside the lab. Near the end of my master's studies I worked in the cell-biology lab of one such scientist. His endless enthusiasm fostered a great lab environment. I accepted an offer to stay for my PhD. I don't regret this decision; my PhD was exciting and successful. But towards the end I struggled with what to do next.

Up to that point, I hadn't had moments of crisis like many PhD students who fear they'll never publish. From the start we formed exciting, successful collaborations. Still, I did not have the passion for cell biology that I saw in my supervisor on a daily basis. Then I started to develop an interest in programming. Along with a friend and fellow PhD student, I developed several scientific applications to help molecular biologists analyse and construct DNA

sequences. It became my passion. Outside the lab, I stopped thinking about cell biology and started thinking about my programs.

What to do next? I considered a hybrid subject, such as electron microscopy, but had difficulty devising a grant proposal. My rescue came in the form of the annual postdoc retreat organized by the Netherlands Cancer Institute. A break from the lab bench and stimulating lectures helped me think about my options. A talk about community-based software that might lead to a more open discussion and valuation of scientific articles sparked lots of research ideas.

Two months ago, I started at the text mining group at the European Bioinformatics Institute in Hinxton near Cambridge, UK. I already feel right at home in this completely different field. I now realize that pursuing some 'hybrid' postdoc would have been a mistake. My creativity and passion lie with computers and bioinformatics. If I'd compromised rather than making the leap to a new field, I would have been unhappy and less enthusiastic. When I discovered I was climbing the wrong ladder I had to make the jump to one that suited me more. ■

Alexander Griekspoor is a Marie Curie postdoctoral fellow at the EMBL European Bioinformatics Institute.

POSTDOC JOURNAL

Smart moves

'Work smarter, not harder.' This is my mantra. I'm a postdoc in a meiotic recombination lab at Cornell University and the mother of 1.5 children. I've found that balancing work and family takes more coordination, organization and time-management skill than I ever needed before starting a family.

Sounds awful, right? Actually it's been the best thing to happen to me. It has made me a better scientist because I carefully plan experiments and maximize output with minimal input. After my son was born, I completed my graduate work, wrote a thesis, graduated, wrote a grant (awarded), wrote a manuscript (accepted), and got a postdoc position. That's not bad for 20 months' work.

With that said, this year brings many new challenges: a new project in an unfamiliar field, several technically difficult project goals and a new baby on the way. Can I still be productive and competitive with all this on my plate? Is the tenured professor track the right career path for me, or should I look to industry?

This year, through conferences and Cornell University's postdoc advisory resources, I plan to talk to women in industry, in faculty and in postdoc positions, and to women scientists who have left science, to hear all sides and garner advice. I'll ask my spouse for more flexibility and support to ease my home life. As for managing research and life, I will rely on my mantra to work smarter, not harder, in all that I do. ■

Moir Sheehan is a postdoc in plant breeding and genetics at Cornell University.

Crossing boundaries, hitting barriers

Interdisciplinary research may be lauded, but it's not yet rewarded.



Adina Payton



Mary Lou Zoback

The scientific community recognizes the importance of interdisciplinary research. But institutions have not yet caught up with ways to reward good examples of it. In fact, interdisciplinary research in academia often faces discouraging barriers. Working across disciplines requires not only depth of knowledge but also a holistic understanding of complex systems. But the tenure system largely favours narrowly focused research in subdisciplines.

Much progress in our field, Earth sciences, has emerged from working across several disciplines. Combining medicine and geophysics is resulting in new medical devices. Mixing meteorology and biology provides new ways of looking at the spread of infectious disease. And blending oceanography and atmospheric science leads to a better understanding of the causes and consequences of global warming. We need to continue and expand these multidimensional approaches to benefit humanity and improve the prediction of climate and natural hazards, and to provide science-based solutions that support human well-being and the sustainable use of our resources.

The 2006 US National Academies report *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* emphasizes the value of interdisciplinary research, noting that many significant scientific and engineering advances cut across several disciplines.

Academia has taken some steps to embrace interdisciplinary research. The Harvard University Center for the Environment programme recognizes that the most pressing problems facing the environment are complex and often require collaborative investigation by scholars versed in different disciplines. Similarly, the new Woods Institute for

the Environment at Stanford University is an interdisciplinary hub for research, teaching and problem-solving that draws on the experience and expertise of faculty members and students from all seven schools at Stanford University.

But the university reward system has generally not kept pace with this approach. Most universities still rely on a tenure process that judges excellence and leadership in a narrowly defined disciplinary (or subdisciplinary) field.

Indeed, a 2004 National Academies report, *Facilitating Interdisciplinary Research*, lists "promotion criteria" as the top impediment to interdisciplinary research, based on separate rankings by both scientists and university provosts. "An interdisciplinary faculty member seeking tenure often faces two challenges beyond those faced by members working in a single discipline," says the report. "First, interdisciplinary research done

"If universities want to attract innovative young researchers, they should support integrative research."

by the candidate may not be valued sufficiently to compensate for lower output of disciplinary research ... Second, it can be difficult to find reviewers who understand the overall quality of the work, which usually lies outside the expertise of people on the tenure evaluation committee."

The committee recommended that institutions provide more flexibility in promotion and tenure procedures, recognizing that the contributions of someone in interdisciplinary research may need to be evaluated differently from those of someone in a single-discipline project. Indeed, if universities want to attract and retain

innovative young researchers and foster novel approaches to science, they should support infrastructure, research needs and opportunities for integrative research. They also need to find different ways of rewarding and evaluating scientists who are engaged in such research.

We advocate several steps towards proper recognition of the contributions of interdisciplinary researchers. They include establishing interdisciplinary review committees to evaluate faculty members who are conducting such research, with at least one of the committee members actively doing interdisciplinary work themselves. We'd like to see tenure committees solicit input from scientists personally familiar with the candidate's work — people who can assess individual contributions to collaborative projects and roles in facilitating the research. Letters to external referees should be formulated to emphasize the transdisciplinary nature of the candidate's work and not require that the candidate be identified as or compared with an expert in one specific field. Interdisciplinary researchers could also have more time to reach tenure milestones, as their research can be more time-consuming to coordinate, conduct and synthesize.

Interdisciplinary research and education are inspired by the drive to provide effective solutions to complex questions. A central problem it faces is in finding ways to remove those barriers. ■

Adina Payton is in the Department of Geological and Environmental Sciences at Stanford University. Mary Lou Zoback is vice-president of earthquake risk applications at Risk Management Solutions in Newark, California.

Facilitating Interdisciplinary Research
 ▶ www.nap.edu/openbook/0309094356/html/R1.html